

Subblock-Constrained Codes for Real-Time Simultaneous Energy and Information Transfer

Anshoo Tandon, *Student Member, IEEE*, Mehul Motani, *Member, IEEE*,
and Lav R. Varshney, *Senior Member, IEEE*

Preprint: Submitted for publication, May 2015.

Abstract

Consider an energy-harvesting receiver that uses the same received signal both for decoding information and for harvesting energy, which is employed to power its circuitry. In the scenario where the receiver has limited battery size, a signal with bursty energy content may cause power outage at the receiver since the battery will drain during intervals with low signal energy. In this paper, we consider a discrete memoryless channel and characterize achievable information rates when the energy content in each codeword is regularized by ensuring that sufficient energy is carried within every subblock duration. In particular, we study constant subblock-composition codes (CSCCs) where all subblocks in every codeword have the same fixed composition, and this subblock-composition is chosen to maximize the rate of information transfer while meeting the energy requirement. Compared to constant composition codes (CCCs), we show that CSCCs incur a rate loss and that the error exponent for CSCCs is also related to the error exponent for CCCs by the same rate loss term. We show that CCCC capacity can be improved by allowing different subblocks to have different composition while still meeting the subblock energy constraint. We provide numerical examples highlighting the tradeoff between delivery of sufficient energy to the receiver and achieving high information transfer rates. It is observed that the ability to use energy in real-time imposes less of penalty than the ability to use information in real-time.

A. Tandon and M. Motani are with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117583 (email: anshoo@nus.edu.sg, motani@nus.edu.sg).

L. R. Varshney is with the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (email: varshney@illinois.edu).

This work was supported in part by the National Research Foundation Singapore under Grant No. NRF-CRP-8-2011-01, and by Systems on Nanoscale Information fabriCs (SONIC), one of the six SRC STARnet Centers, sponsored by MARCO and DARPA.

Some results in this paper were presented in part at the IEEE SECON 2014 Workshop on Energy Harvesting Communications, June, 2014 [1], and at the 2015 International Symposium on Information Theory (ISIT), June, 2015 [2].

I. INTRODUCTION

Although wireless charging of portable electronic devices [3] and implantable biomedical devices [4] has attracted the attention of researchers over the last few years, pioneering work on wireless power transfer was conducted over a century ago by Hertz and Tesla [5]. Similarly, wireless information transfer has a rich history, including works by Popov [6], Bose [7], and Marconi [8]. In fact, Marconi's wireless telegraph device, capable of transatlantic radio communication, helped save over 700 lives during the tragic accident of the Titanic in 1912 [9]. However, the first work in an information-theoretic setting on analyzing fundamental tradeoffs between *simultaneous* information and energy transfer is relatively recent [10]. The study of simultaneous information and energy transfer is relevant for communication from a powered transmitter to an energy-harvesting receiver which uses the same received signal both for decoding information and for harvesting energy. The energy harvested by the receiver is employed to power its circuitry.

The tradeoff between reliable communication and delivery of energy at the receiver was characterized in [10] using a general capacity-power function, where transmitted codewords were constrained to have average received energy exceed a threshold. This tradeoff between capacity and energy delivery was extended for frequency-selective channels in [11]. Since then, there have been numerous extensions of the capacity-power function in various settings [12]–[15]. Biomedical applications of wireless energy and information transfer have been proposed through the use of implanted brain-machine interfaces that receive data and energy through inductive coupling [4], [16], [17].

However, in practical applications such as biomedical, imposing only an average power constraint is not sufficient; we also need to regularize the transferred energy content. This is because a codeword satisfying the average power constraint may still cause outage at the receiver if the energy content in the codeword is bursty, since the receive energy buffer with a relatively small storage capacity may drain during intervals with low signal energy. In order to regularize the energy content in the signal, we herein adopt a *subblock-constrained* approach where codewords are divided into smaller subblocks, and every subblock is constrained to carry sufficient energy exceeding a given threshold. The subblock length and the energy threshold may be chosen to meet the real-time energy requirement at the receiver.

An alternative to the subblock-constraint is the sliding-window constraint, which we do not consider here. Under a sliding-window constraint, each codeword provides sufficient energy

within a sliding time window of certain duration. This approach was adopted in [18], [19], where the use of runlength codes for simultaneous energy and information transfer was proposed. In [20], a sliding window constraint was imposed on binary codewords and bounds on the capacity were presented for different binary input channels. Note that the sliding-window constraint is relatively tighter than the subblock-constraint, since subblock-constraint corresponds to the case where the windows are non-overlapping.

In this paper, we consider a discrete memoryless channel (DMC) and characterize achievable information rates when each *subblock* is constrained to carry sufficient energy. We assume that corresponding to transmission of each symbol in the input alphabet, the receiver harvests a certain amount of energy as a function of the transmitted symbol. Since different symbols may correspond to different energy levels, the requirement of sufficient energy content within a subblock imposes a constraint on the composition of each subblock. Towards meeting this subblock energy requirement, we introduce the *constant subblock-composition codes* (CSCCs) where all the subblocks in every codeword have the same fixed composition. This subblock-composition, quantifying the fraction of different symbols with each subblock, is chosen to maximize the rate of information transfer while meeting the energy requirement. Note that if x_1^L denotes a given subblock of length L , then the composition of x_1^L is the distribution $P_{x_1^L}$ on \mathcal{X} defined by $P_{x_1^L}(x) \triangleq \frac{N(x)}{L}$, $x \in \mathcal{X}$, where $N(x)$ is the number of occurrences of symbol x in subblock x_1^L .

A. Our Contribution

For meeting the real-time energy requirement at a receiver which uses the received signal to simultaneously harvest energy and decode information, we propose the use of CSCCs (Sec. III-A) and establish their capacity as a function of the required energy per symbol (Sec. III-B). We show that CSCC capacity can be computed efficiently by exploiting certain symmetry properties (Sec. III-C) and present bounds on subblock length for avoiding receiver energy outage (Sec. III-D).

Compared to constant composition codes, we quantify the rate loss incurred due to the additional constraint of restricting all subblocks within codewords to have the same composition (Sec. IV-A). For a given rate of information transfer, we derive a lower bound for the error exponent using CSCC in terms of the error exponent for constant composition codes (Sec. IV-B).

We show that information rates greater than CSCC capacity can be achieved by allowing different subblocks to have different composition, while still meeting the energy requirement per subblock (Sec. V).

For enabling real-time information transfer, we consider local subblock decoding where each subblock is decoded independently (Sec. VI), and compare achievable rates using local subblock decoding with those when all the subblocks within a codeword are jointly decoded. We also provide numerical results highlighting the tradeoff between delivery of sufficient energy to the receiver and achieving high information rates (Sec. VII).

B. Related Work

Codes with different constraints on the codewords have been suggested in the past, depending on the constraints at the transmitter, the properties of the communication channel, or the properties of the storage medium. For digital information storage on magnetic medium [21], codewords are usually designed to meet the runlength constraint [22] or are optimized for partial response equalization with maximum-likelihood sequence detection (PRML) [23]. The study of information capacity using runlength-limited (RLL) codes on binary symmetric channels (BSC) was carried in [24]–[26].

A class of binary block codes called *multiply constant-weight codes* (MCWC), where each codeword of length mn is partitioned into m equal parts and has weight w in each part, was explored in [27] owing to their potential application in implementation of low-cost authentication methods [28]. Note that MCWC, introduced in [27] as a generalization of *constant weight codes* [29], are themselves a special case of CSCCs with input alphabet size equal to two. When each codeword in an MCWC is arranged as an $m \times n$ array and additional weight constraints are imposed on all the columns, the resulting two-dimensional weight constrained codes have potential application in optical storage systems [30] and in power line communications [31].

Power line communications (PLC) requires the power output to be as constant as possible so that information transfer does not interfere with the primary function of power delivery. One way to achieve this on the PLC channel (which suffers from narrow-band interference, white Gaussian noise, and impulse noise [32]), is to employ *permutation codes* [33] where each codeword of length n is a permutation of n different frequencies, with each frequency viewed as an input symbol. Higher rates of information transfer may be achieved using *constant composition codes* [34] at the cost of local variation in power while ensuring that the power

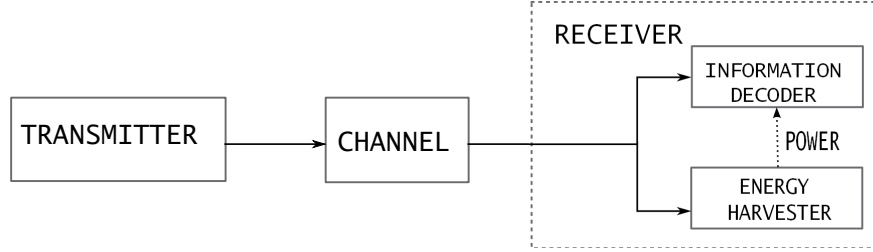


Fig. 1. Simultaneous information and energy transfer from a transmitter to an energy-harvesting receiver

expended is same upon completion of each codeword. When the codeword length is a multiple of the frequency alphabet size, the composition may be chosen such that each frequency occurs equal number times in each codeword [35].

The codewords employed by an energy harvesting transmitter are constrained by the instantaneous energy available for transmission. The capacity of these constrained codes over an additive white Gaussian noise (AWGN) channel has been analyzed when the energy storage capability at the transmitter is zero [36], infinite [37], or some finite quantity [38], [39]. The capacity of an AWGN channel with processing cost at an energy harvesting transmitter was characterized in [40]. The DMC capacity using an energy harvesting transmitter equipped with a finite energy buffer was analyzed in [41]. A comprehensive summary of the recent contributions in the broad area of energy harvesting wireless communications was provided in [42].

II. SYSTEM MODEL

Consider communication from a transmitter to a receiver where the receiver uses the received signal both for decoding information as well as for harvesting energy (see Fig. 1). We model the effective communication channel from the output of a digital modulator at the transmitter to the input to an information decoder at the receiver as a DMC. Note that a DMC is characterized by input alphabet \mathcal{X} , output alphabet \mathcal{Y} , and a stochastic matrix $W : \mathcal{X} \rightarrow \mathcal{Y}$ with $W = \{W(y|x) : x \in \mathcal{X}, y \in \mathcal{Y}\}$ where the matrix entry $W(y|x)$ is the probability that the output is y when the channel input is x .

A DMC is a reasonable communication channel model for simultaneous energy and information transfer. Consider, for instance, the use of a digital modulator at the transmitter which produces symbols from a signal constellation $\mathcal{X} = \{x_1, \dots, x_r\}$. At the receiver, the signal is split for use by the energy harvesting module and the information processing module, respectively.

The input to the information decoder at the receiver comprises of one of s quantized values $\mathcal{Y} = \{y_1, \dots, y_s\}$, fed by a quantizer in the information processing path. For each quantized value $y_i, 1 \leq i \leq s$, and each transmitted symbol $x_j, 1 \leq j \leq r$, the likelihood $\Pr(y_i|x_j)$ can be computed based on the effective signal path from the transmit modulator to the quantizer at the receiver. The communication channel is thus a DMC with input alphabet \mathcal{X} , output alphabet \mathcal{Y} , and channel transition probabilities $\Pr(y_i|x_j)$.

In practice, the effective channels seen by the information decoder and the energy harvester may be different due to their respective pre-processing stages. A simple time-sharing approach to transmitting energy and information simultaneously was suggested in [43] via interleaving of energy signal and information-bearing signal. In [44], practical architectures for simultaneous information and energy reception were defined: an “integrated” receiver architecture has shared radio frequency chains between the energy harvester and the information decoder, whereas a “separated” architecture has different chains.

In our work, we assume a generic receiver architecture where the received signal is split between the energy harvesting path and the information processing path with a static power splitting ratio. The effective communication channel seen by the decoder in the information processing path is modeled as a DMC. We let $b(x)$ denote the energy harvested by the harvester after the signal split at the receiver, when $x \in \mathcal{X}$ is transmitted. Thus, b is a map from the input alphabet \mathcal{X} to the set of non-negative real numbers, and higher energy is carried by symbols having higher b -value. This map is assumed to be time-invariant, and reflects the scenario where the statistical nature of the effective communication channel is due to the noise in the receiver circuitry, which does not affect the harvested energy. The quantification of b abstracts the specific implementation of a chosen receiver architecture, which in turn helps to abstract the problem of the code design for simultaneous energy and information transfer from implementation details.

In order to meet the real-time energy requirement at the receiver, we partition the transmitted codeword into equal-sized subblocks (see Fig. 2) and require that transmitted symbols be chosen such that the expected harvested energy in each subblock exceeds a given threshold. This threshold is a function of the energy consumption by the receiver circuitry including the information decoder. We will denote the subblock length by L and assume that the codeword length, denoted n , is a multiple of L . If a transmitted codeword is denoted (X_1, X_2, \dots, X_n) ,

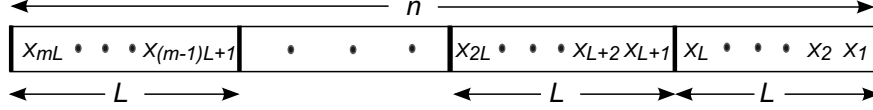


Fig. 2. Transmitted codeword partitioned into subblocks of length L .

then the constraint on sufficient energy within each subblock can be expressed as

$$\frac{1}{L} \sum_{i=1}^L b(X_{(j-1)L+i}) \geq B, \quad j = 1, 2, \dots, m \quad (1)$$

where j is the subblock index, B denotes the required energy per symbol at the receiver, and m is the number of subblocks in a codeword. The choice of the subblock length L depends on the energy storage capacity at the receiver; a small energy buffer generally requires relatively small value of L to prevent energy outage at the receiver.

The subblock energy constraint given by (1) becomes trivial if $b(x)$ is same for all $x \in \mathcal{X}$ (for instance, when the transmitted symbols belong to a phase-shift-keying constellation). However, the constraint is non-trivial when b -values are not constant (for instance, using on-off keying) and threshold B satisfies

$$b_{\min} < B < b_{\max}, \quad (2)$$

where

$$b_{\min} = \min_{x \in \mathcal{X}} b(x), \quad b_{\max} = \max_{x \in \mathcal{X}} b(x). \quad (3)$$

In the rest of the paper we assume (2) is satisfied, unless otherwise stated.

For a given subblock j within a codeword, if $N(x)$ denotes the number of occurrences of symbol x in the j th subblock, then (1) can alternately be expressed as

$$\sum_{x \in \mathcal{X}} b(x) \frac{N(x)}{L} \geq B. \quad (4)$$

Note that $N(x)/L$ denotes the fraction of time when symbol x appears in the subblock. We now introduce constant subblock-composition codes which are a nice way to meet the subblock energy constraint.

III. CONSTANT SUBBLOCK-COMPOSITION CODES

A. Motivation and Definition

We have seen that for a given subblock, the energy constraint given by (1) can equivalently be expressed as (4) and this constraint is satisfied provided the fraction of time each symbol

appears in the subblock is chosen appropriately. This observation motivates the use of codes where the composition of each subblock in all codewords is constant and is chosen such that (4) is satisfied. A *constant subblock-composition code* (CSCC) is one in which all codewords are partitioned into equal-sized subblocks and each subblock (in all codewords) has the same type P . The subblock type P in CSCC is chosen to satisfy the subblock energy constraint

$$\mathbb{E}_P[b(X)] \triangleq \sum_{x \in \mathcal{X}} b(x)P(x) = \sum_{x \in \mathcal{X}} b(x) \frac{N(x)}{L} \geq B. \quad (5)$$

B. Capacity using CSCC

Let \mathcal{P}_L denote the set of all compositions for input sequences of length L . For a given type $P \in \mathcal{P}_L$, the set of sequences in \mathcal{X}^L with composition P is denoted by \mathcal{T}_P^L and is called the *type class* or *composition class* of P . In a CSCC with subblock-composition P , every subblock in a codeword may be viewed as an element of \mathcal{T}_P^L .

In order to compute the capacity of a CSCC on a DMC, we may view the L uses of the original channel as a single use of the induced *vector channel* having input alphabet \mathcal{T}_P^L and output alphabet \mathcal{Y}^L . Since the underlying channel is memoryless, the transition probabilities for a pair of input and output vectors is the product of the corresponding transition probabilities of the underlying channel. If we let $x_1^L = x_1 \dots x_L$ and $y_1^L = y_1 \dots y_L$ be given input and output vectors with $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$, respectively, then the transition probabilities for the induced vector channel are:

$$W^L(y_1^L|x_1^L) = \prod_{i=1}^L W(y_i|x_i). \quad (6)$$

Since each subblock in a codeword may be chosen independently, the capacity using CSCC with subblock-composition P , denoted $C_{CSCC}^L(P)$, is equal to $1/L$ times the capacity of the induced vector channel with input alphabet \mathcal{T}_P^L , output alphabet \mathcal{Y}^L , and transition probabilities given by (6). Thus if we denote $X_1^L = X_1 \dots X_L$ and $Y_1^L = Y_1 \dots Y_L$, then

$$C_{CSCC}^L(P) = \max_{X_1^L \in \mathcal{T}_P^L} \frac{I(X_1^L; Y_1^L)}{L} \quad (7)$$

$$= \max_{X_1^L \in \mathcal{T}_P^L} \left(\frac{H(Y_1^L)}{L} - \frac{H(Y_1^L|X_1^L)}{L} \right) \quad (8)$$

$$= \max_{X_1^L \in \mathcal{T}_P^L} \left(\frac{H(Y_1^L)}{L} - \frac{\sum_{i=1}^L H(Y_i|X_i)}{L} \right) \quad (9)$$

where the last equality follows from the memoryless property of the channel. The maximization in (7) is over the distribution of input vector in \mathcal{T}_P^L . We will show that the maximum is achieved when the input vectors X_1^L are uniformly distributed over \mathcal{T}_P^L .

Theorem 1. *The capacity of the induced vector-channel using CSCC with fixed subblock-composition P is obtained via a uniform distribution of the input vectors in \mathcal{T}_P^L .*

Proof: See Appendix A. ■

If we define the set of distributions

$$\Gamma_B^L \triangleq \{P \in \mathcal{P}_L : \mathbb{E}_P[b(X)] \geq B\}, \quad (10)$$

then the capacity using CSCC with subblock energy constraint (1), denoted $C_{CSCC}^L(B)$, is defined as

$$C_{CSCC}^L(B) = \max_{P \in \Gamma_B^L} C_{CSCC}^L(P) \quad (11)$$

C. Computing CSCC Capacity

By Theorem 1, the maximum is achieved in (9) when X_1^L is uniformly distributed over \mathcal{T}_P^L . The computation of the capacity expression with increasing subblock length L seems challenging since the input and output alphabet size for the induced vector channel grows exponentially with L . However, we will show that the computational complexity of the CSCC capacity expression can be reduced using the following observations.

First note that the probability distribution for the output vector in the induced vector channel is given by

$$P_{Y_1^L}(y_1^L) = \frac{1}{|\mathcal{T}_P^L|} \sum_{x_1^L \in \mathcal{T}_P^L} W^L(y_1^L | x_1^L), \quad (12)$$

since the input vectors are uniformly distributed over \mathcal{T}_P^L . If \tilde{y}_1^L is another output vector having the same composition as y_1^L , then we have $P_{Y_1^L}(y_1^L) = P_{Y_1^L}(\tilde{y}_1^L)$. This is because the columns $W^L(y_1^L | \cdot)$ and $W^L(\tilde{y}_1^L | \cdot)$ of the vector channel transition matrix are permutations of each other (see Appendix A). Thus output vectors having the same composition have equal probability. However, even though the input vectors are uniformly distributed, the output vectors in general are *not* uniformly distributed. Also, since the symbols within an input vector $x_1^L \in \mathcal{T}_P^L$ are not independent, in general we have $P_{Y_1^L}(y_1^L) \neq \prod_{i=1}^L P_Y(y_i)$, where $P_Y(y)$ denotes the probability of output scalar symbol y .

Let \mathcal{Q}_L denote the set of all compositions for output sequences of length L . When X_1^L is uniformly distributed over \mathcal{T}_P^L , the $H(Y_1^L)$ term in (9) can be expressed as

$$H(Y_1^L) = - \sum_{y_1^L \in \mathcal{Y}^L} P_{Y_1^L}(y_1^L) \log P_{Y_1^L}(y_1^L) \quad (13)$$

$$= - \sum_{Q \in \mathcal{Q}_L} \sum_{y_1^L \in \mathcal{T}_Q^L} P_{Y_1^L}(y_1^L) \log P_{Y_1^L}(y_1^L) \quad (14)$$

$$= \sum_{Q \in \mathcal{Q}_L} |\mathcal{T}_Q^L| P_{Y_1^L}(y_1^L) \log \frac{1}{P_{Y_1^L}(y_1^L)}, \quad (15)$$

where the last equality follows because $P_{Y_1^L}(y_1^L)$ is same for all $y_1^L \in \mathcal{T}_Q^L$. Note that we choose only one representative vector y_1^L from each type class \mathcal{T}_Q^L in the last equality.

Secondly, the following proposition shows that the $H(Y_i|X_i)$ term in (9) is same for all $1 \leq i \leq L$, since the corresponding joint probabilities $P_{XY}(X_i = x, Y_i = y)$ are equal.

Proposition 1. *For a random input vector X_1^L uniformly distributed over \mathcal{T}_P^L with corresponding output vector Y_1^L , the pairwise probability $P_{XY}(X_i = x, Y_i = y)$, for $1 \leq i \leq L$, satisfies*

$$P_{XY}(X_i = x, Y_i = y) = \frac{N(x)}{L} W(y|x) = P(x)W(y|x). \quad (16)$$

Proof: Since

$$P_{XY}(X_i = x, Y_i = y) = \Pr(X_i = x)W(y|x), \quad (17)$$

the claim will be proved if we show $\Pr(X_i = x) = N(x)/L$ for all $1 \leq i \leq L$. As X_1^L is uniformly distributed over \mathcal{T}_P^L , the $\Pr(X_i = x)$ is equal to the ratio of the number of input vectors with x at index i to the total number of vectors in \mathcal{T}_P^L . Since

$$|\mathcal{T}_P^L| = \frac{L!}{\prod_{x \in \mathcal{X}} N(x)!}, \quad (18)$$

and the number of sequences in \mathcal{T}_P^L with x at index i is

$$\frac{(L-1)!}{(N(x)-1)! \prod_{\tilde{x} \neq x} N(\tilde{x})!}, \quad (19)$$

the ratio of the quantities given by (19) and (18) is equal to $\Pr(X_i = x) = N(x)/L$. ■

The next proposition gives a computationally efficient expression for CSCC capacity.

Proposition 2. *The CSCC capacity, $C_{CSCC}^L(B)$, is given by*

$$\max_{P \in \Gamma_B^L} \frac{1}{L} \sum_{Q \in \mathcal{Q}_L} |\mathcal{T}_Q^L| P_{Y_1^L}(y_1^L) \log \frac{1}{P_{Y_1^L}(y_1^L)} - H(Y|X), \quad (20)$$

where only one representative output vector y_1^L is chosen from every type class \mathcal{T}_Q^L , $P_{Y_1^L}(y_1^L)$ is given by (12), and $H(Y|X)$ is evaluated using the joint pairwise probability distribution given by (16).

Proof: Use (9) and (11) to express $C_{CSCC}^L(B)$. From Thm. 1, a uniform distribution over \mathcal{T}_P^L achieves capacity, and hence the entropy term $H(Y_1^L)$ in (9) can be computed using (15). The claim in Prop. 2 follows by further noting that the $H(Y_i|X_i)$ term in (9) is the same for all $1 \leq i \leq L$, which can be evaluated using the joint pairwise distribution in (16). ■

D. Choice of Subblock Length L

In this subsection, we derive bounds on subblock length L (as a function of the energy storage capacity at the receiver) which will ensure that the receiver never runs out of energy when the subblock-composition P is chosen to satisfy (5). It will be seen that a large energy storage capacity allows for larger values of L and hence results in higher rates of information transfer.

The energy storage capacity at the receiver is denoted E_{max} and we assume that the receiver requires B units of energy per received symbol for its processing. Let $E(i)$ denote the level of the energy buffer at the receiver at the completion of $i - 1$ uses of the channel. The energy update equation, for $i = 1, 2, \dots$, is given by

$$E(i + 1) = \min(E_{max}, |E(i) + b(X_i) - B|^+), \quad (21)$$

where X_i is the symbol transmitted in the i th channel use, and $|z|^+ \triangleq \max(z, 0)$.

We say that an *outage* occurs during i th channel use if

$$E(i) + b(X_i) < B, \quad (22)$$

while an *overflow* event occurs if

$$E(i) + b(X_i) - B > E_{max}. \quad (23)$$

We partition the input alphabet as $\mathcal{X} = \mathcal{X}_\triangleleft \cup \mathcal{X}_\triangleright$, where

$$\mathcal{X}_\triangleleft = \{x \in \mathcal{X} \mid b(x) < B\}, \quad (24)$$

$$\mathcal{X}_\triangleright = \{x \in \mathcal{X} \mid b(x) \geq B\}. \quad (25)$$

For CSCC with subblock-composition $P \in \Gamma_B^L$, we define

$$G = \sum_{x \in \mathcal{X}_\triangleleft} LP(x) (B - b(x)), \quad (26)$$

where G will be used to characterizes some useful properties of the energy update process.

Lemma 1. *The energy update process satisfies the following properties for CCCC with subblock-composition $P \in \Gamma_B^L$:*

- (a) *If there is no energy outage or overflow during the reception of the first subblock, then*

$$E(L+1) \geq E(1).$$
- (b) *If $E(1) \geq G$, then there is no energy outage during the reception of the first subblock.*
- (c) *If $E(1) \geq G$ and $E_{max} \geq 2G$, then $E(L+1) \geq G$.*

Proof: If there is no energy outage or overflow, then the total energy harvested during the reception of the first subblock is $\sum_{x \in \mathcal{X}} LP(x)b(x)$, while the total energy consumed is LB and claim (a) follows since P satisfies (5).

Let X_i denote the transmitted symbol in the i th channel use, $I = \{1, 2, \dots, L\}$, and $I_{<} = \{i \in I | X_i \in \mathcal{X}_{<}\}$. For $i \in I$, the level in the energy buffer decreases during the i th channel use if and only if $i \in I_{<}$, and the corresponding decrease in energy level is $B - b(X_i)$. Since the subblock has composition P , the sum of energy decrements over the reception of the first subblock is $\sum_{i \in I_{<}} B - b(X_i) = G$, and claim (b) follows.

For proving claim (c), we note that the condition $E(1) \geq G$ implies that there is no energy outage during the reception of the first subblock (using claim (b)). Further, if there is no overflow then $E(L+1) \geq E(1) \geq G$ (using claim (a)). In case there is energy overflow in the i th channel use for any $i \in I$, we have $E(i+1) = E_{max} \geq 2G$, and thus $E(L+1) \geq E(i+1) - G \geq G$. ■

Lemma 1 is useful in proving the following theorem which gives a necessary and sufficient condition on subblock length in order to avoid outage.

Theorem 2. *A necessary and sufficient condition on L for avoiding energy outage during the reception of CCCC codewords, with subblock-composition P satisfying (5), is*

$$L \leq \frac{E_{max}}{\sum_{x \in \mathcal{X}_{<}} 2P(x)(B - b(x))}, \quad (27)$$

with $E(1) \geq G$.

Proof: See Appendix B. ■

The initial condition on energy level, $E(1) \geq G$, may be ensured by transmitting a preamble, consisting of symbols with high energy content, before the transmission of codewords. This preamble has bounded length and hence does not affect the channel capacity.

IV. COMPARING CSCC WITH CONSTANT COMPOSITION CODES

A. Rate Comparison

Similar to subblock-composition, a codeword composition represents the fraction of times each input symbol occurs in a codeword and a constant composition code (CCC) is one in which all codewords have the same composition. Note that a CCCC with subblock-composition P may also be viewed as a CCC with codeword composition P , since all the subblocks in CCCC have the same composition. In general for CCC, although all codewords have the same composition, different subblocks within a codeword may have different compositions. Hence CCCs are richer than CCCCCs in terms of choice of symbols within each subblock. CCCs were first analyzed by Fano [45] and shown to be sufficient to achieve capacity for any discrete memoryless channel.

Let $C_{CCC}(P)$ denote the maximum achievable rate using CCC with codeword composition P . For $P \in \Gamma_B^L$ (refer (10)), a CCC with *codeword* composition P will ensure that the average received energy per symbol in a codeword is at least B . However, it may violate the constraint on providing sufficient energy to the receiver *within every subblock duration*. For a CCC, we have [45]

$$C_{CCC}(P) = I(X; Y) = H(X) - H(X|Y). \quad (28)$$

We are interested in quantifying the information rate penalty incurred by using CCCC compared to CCC, given by $C_{CCC}(P) - C_{CSCC}^L(P)$. This information rate penalty is the price we pay for meeting the real-time energy requirement within every subblock duration, compared to the less constrained energy requirement per codeword. Although the rate penalty can be numerically computed by explicit computation of $C_{CCC}(P)$ and $C_{CSCC}^L(P)$, the numerical approach has the limitation that the computation complexity of $C_{CSCC}^L(P)$ increases with an increase in subblock L .

In CCCC, since a transmitted subblock X_1^L is uniformly distributed over \mathcal{T}_P^L , we have [46, p. 26]

$$H(X_1^L) = \log |\mathcal{T}_P^L| = LH(P) - Lr(L, P), \quad (29)$$

where $r(L, P)$ denotes a function of L and P given as

$$r(L, P) = \frac{s(P) - 1}{2L} \log(2\pi L) + \frac{1}{2L} \sum_{a: P(a) > 0} \log P(a) + \frac{\vartheta(L, P)}{12L \ln 2} s(P), \quad (30)$$

with $s(P)$ denoting the number of elements $x \in \mathcal{X}$ with $P(x) > 0$, and $\vartheta(L, P)$ is a real number between zero and one which is chosen so that (29) is satisfied.

We now present simple analytical bounds for this rate penalty. The following theorem shows that the rate penalty by using CSCC, relative to CCC, is bounded by $r(L, P)$.

Theorem 3. *The rate penalty is bounded as*

$$0 \leq C_{CCC}(P) - C_{CSCC}^L(P) \leq r(L, P). \quad (31)$$

Further, there exist channels for which the rate penalty meets the upper or lower bound in (31) with equality.

Proof: When X_1^L is uniformly distributed over \mathcal{T}_P^L ,

$$C_{CSCC}^L(P) = \frac{1}{L} [H(X_1^L) - H(X_1^L|Y_1^L)] \quad (32)$$

$$\stackrel{(a)}{=} H(P) - r(L, P) - \frac{1}{L} \sum_{i=1}^L H(X_i|Y_1^L, X_1^{i-1})$$

$$\stackrel{(b)}{\geq} H(P) - r(L, P) - \frac{1}{L} \sum_{i=1}^L H(X_i|Y_i)$$

$$\stackrel{(c)}{=} H(P) - r(L, P) - H(X|Y)$$

$$\stackrel{(d)}{=} C_{CCC}(P) - r(L, P), \quad (33)$$

where X_1^{i-1} denotes X_1, \dots, X_{i-1} , (a) follows from (29) and chain rule for entropy, (b) follows since conditioning only reduces entropy, (c) follows from (16), and (d) follows from (28). Now, (31) follows from (33). Explicit channels can be constructed which meet the bounds in (31).

- $C_{CCC}(P) = C_{CSCC}^L(P) = 0$ for a binary symmetric channel (BSC) with crossover probability equal to 0.5.
- For a noiseless channel, we have $C_{CCC}(P) - C_{CSCC}^L(P) = r(L, P)$ due to equality in (b) as $\sum_{i=1}^L H(X_i|Y_1^L, X_1, \dots, X_{i-1}) = \sum_{i=1}^L H(X_i|Y_i) = 0$.

■

Corollary 1.

$$\lim_{L \rightarrow \infty} C_{CSCC}^L(P) = C_{CCC}(P) \quad (34)$$

Proof: Note that for a fixed P , the value of $r(L, P)$ as a function of L is non-negative and falls roughly as $\log(L)/L$ and thus tends to zero as $L \rightarrow \infty$. Thus (34) follows by taking the limit $L \rightarrow \infty$ in (31).

■

Remark: For a fixed subblock length L , the CCCC capacity can be achieved by making the number of subblocks in a codeword arbitrarily large and performing joint decoding over all the subblocks. However, when the number of subblocks in a codeword are kept constant and the subblock length is increased without bounds, then achievable rates using CCCC tend to CCC capacity. In particular, when there is only one subblock in a codeword, then the CCCC code is same as a CCC code whose capacity can be achieved by making L arbitrarily large.

The upper bound (30) on the rate penalty given by $r(L, P)$ is independent of the underlying channel. In general, given a communication channel, the bounds on rate penalty can be further improved. Consider, for example, a BSC with crossover probability p_0 where $0 < p_0 < 0.5$. For this channel, the upper bound can be tightened using Thm. 4. We first define a binary operator \star and a function h , respectively, as

$$a \star b \triangleq a(1 - b) + (1 - a)b. \quad (35)$$

$$h(x) \triangleq -x \log x - (1 - x) \log(1 - x). \quad (36)$$

We employ the above definitions to state the following theorem on bounding the rate penalty for a BSC.

Theorem 4. *For a BSC with crossover probability $0 < p_0 < 0.5$, input distribution denoted by $P(0) = \Pr(X = 0)$, $P(1) = \Pr(X = 1)$, and $0 < \gamma = \min(P(0), P(1)) \leq 0.5$ we have,*

$$0 < C_{CCC}(P) - C_{CCCC}^L(P) \leq h(p_0 \star \gamma) - h(p_0 \star \alpha) < r(L, P), \quad (37)$$

where α is chosen such that

$$h(\alpha) = h(\gamma) - r(L, P), \quad 0 \leq \alpha < 0.5. \quad (38)$$

Proof: See Appendix C. ■

The proof of Theorem 4 uses Mrs. Gerber's Lemma (MGL) [47]. Using an extension [48] of MGL, the upper bound on the rate penalty can similarly be improved for general memoryless binary-input symmetric-output channels. In particular, we have the following theorem for the binary erasure channel (BEC).

Theorem 5. *For a BEC with erasure probability $\epsilon > 0$,*

$$C_{CCC}(P) - C_{CCCC}^L(P) \leq (1 - \epsilon)r(L, P) < r(L, P) \quad (39)$$

Proof: See Appendix D. ■

For memoryless asymmetric binary-input, binary-output channels, an alternate upper bound on the rate penalty (other than (31)) may be obtained using the equality of the channel *characteristic function* and the *gerbator* [49]. As an example, we have the following theorem for the Z -channel.

Theorem 6. *For a Z -channel with $\gamma = \Pr(X = 1)$, and $p_0 = \Pr(1 \rightarrow 0)$, we have*

$$C_{CCC}(P) - C_{CSCC}^L(P) \leq h(\gamma(1 - p_0)) - h(\alpha(1 - p_0)), \quad (40)$$

where $h(\cdot)$ is given by (36), and α is chosen such that

$$h(\alpha) = h(\gamma) - r(L, P), \quad 0 \leq \alpha < 0.5. \quad (41)$$

Proof: See Appendix E. ■

The rate penalty bound given by (40) may sometimes be worse than the bound in (31), depending on γ and p_0 . In general, the rate penalty for the Z -channel can be upper bounded by $\min(r(L, P), h(\gamma(1 - p_0)) - h(\alpha(1 - p_0)))$.

B. Error Exponent Comparison

In this subsection, we discuss the error exponent using CSCC and show that it can be bounded as a function of the (computationally simpler) error exponent for CCC.

We now present some definitions and notations which will be used in this subsection. For a pair of random variables (X, Y) with $P_X = P$, and conditional probability distribution $P_{Y|X} = W$, we will write $H(Y|X)$ as $H(W|P)$, $I(X; Y)$ as $I(P, W)$, and the distribution of Y as PW . Thus we have

$$PW(y) \triangleq \sum_{x \in \mathcal{X}} P(x)W(y|x), \quad y \in \mathcal{Y} \quad (42)$$

$$H(W|P) \triangleq \sum_{x \in \mathcal{X}} P(x)H(W(\cdot|x)) \quad (43)$$

$$I(P, W) \triangleq H(PW) - H(W|P). \quad (44)$$

The *informational divergence* of distributions P and Q is denoted as

$$D(P||Q) \triangleq \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}. \quad (45)$$

The *conditional informational divergence* of stochastic matrices $V : \mathcal{X} \rightarrow \mathcal{Y}$ and $W : \mathcal{X} \rightarrow \mathcal{Y}$ with respect to distribution P on X is denoted as

$$D(V||W|P) \triangleq \sum_{x \in \mathcal{X}} P(x) D(V(\cdot|x)||W(\cdot|x)) . \quad (46)$$

For CCCs with codeword composition P and information rate $R > 0$, the *sphere packing exponent function* [46] of DMC W is given by

$$E_{sp}(R, P, W) \triangleq \min_{V: I(P, V) \leq R} D(V||W|P) , \quad (47)$$

with V ranging over all channels $V : \mathcal{X} \rightarrow \mathcal{Y}$, and represents an upper bound on the error exponent using best possible codes. For fixed P and W , the function $E_{sp}(R, P, W)$ is a convex function of $R > 0$ (which follows from convexity of $D(V||W|P)$ and $I(P, V)$ as a function of V), positive for $R < I(P, W)$ and zero otherwise.

The *random coding exponent function* [46] of channel W for CCCs with codeword composition P and information rate $R > 0$ is denoted by $E_r(R, P, W)$ and represents a lower bound on achievable error exponent. It is related to $E_{sp}(R, P, W)$ as

$$E_r(R, P, W) = \begin{cases} E_{sp}(R, P, W), & \text{if } R \geq \hat{R} \\ E_{sp}(\hat{R}, P, W) + \hat{R} - R, & \text{if } 0 < R < \hat{R}, \end{cases} \quad (48)$$

where \hat{R} is the smallest R at which the convex curve $E_{sp}(R, P, W)$ meets its supporting line of slope -1 .

The structure of V which achieves the minimum in (47) for $R < I(P, W)$ is given by the following lemma. For $R \geq I(P, W)$, the minimum in (47) is equal to zero which is obtained by choosing $V = W$.

Lemma 2. *For $R < I(P, W)$, the stochastic matrix $V : \mathcal{X} \rightarrow \mathcal{Y}$ which minimizes $D(V||W|P)$ subject to $I(P, V) \leq R$ is given by*

$$V(y|x) = \frac{W(y|x)^{1-s} PV(y)^s}{\sum_{\tilde{y} \in \mathcal{Y}} W(\tilde{y}|x)^{1-s} PV(\tilde{y})^s}, \quad (49)$$

where $PV(y)$ satisfies the set of simultaneous equations

$$PV(y) = \sum_{x \in \mathcal{X}} P(x) V(y|x) = \sum_{x \in \mathcal{X}} \frac{P(x) W(y|x)^{1-s} PV(y)^s}{\sum_{\tilde{y} \in \mathcal{Y}} W(\tilde{y}|x)^{1-s} PV(\tilde{y})^s}, \quad (50)$$

and $s \in [0, 1]$ is chosen such that $I(P, V) = R$.

Proof: See Appendix F. ■

We remark that the random coding exponent function for a DMC was stated by Fano [45] using the distributions $V(y|x)$ and $PV(y)$, given by (49) and (50), respectively, and were referred to as *tilted probability distributions*. However, the explicit statement of Lemma 2 seems not to have appeared in the literature before.

The following theorem uses Shannon's random coding argument to bound the probability of error for CSCC with subblock-composition P on a DMC. It also applies Lemma 2 to compactly express the error probability in terms of the sphere packing exponent function.

Theorem 7. *There exists a CSCC with subblock length L , subblock-composition P , and codeword length n , transmitting information at rate $R > 0$ on DMC W , for which the maximum probability of error is upper bounded as*

$$P_e < \begin{cases} 2 \exp(-n E_{sp}(R', P, W)), & \text{if } R' \geq \hat{R} \\ \exp\left(-n \left(E_{sp}(\hat{R}, P, W) + \hat{R} - R'\right)\right), & \text{if } R' < \hat{R}, \end{cases} \quad (51)$$

where $R' = R + r(L, P)$ and \hat{R} is the smallest R' at which the convex curve $E_{sp}(R', P, W)$ meets its supporting line of slope -1 .

Proof: See Appendix G. ■

The following corollary is immediate.

Corollary 2. *The error exponent for CSCC with subblock length L , subblock-composition P , information rate $R > 0$ on DMC W , is lower bounded by*

$$E_r(R + r(L, P), P, W). \quad (52)$$

Thus the bound on the error exponent for CSCC is related to the error exponent for CCC by the same term, $r(L, P)$, as the bound for the rate penalty (31).

V. BEYOND CONSTANT SUBBLOCK COMPOSITION CODES

In a CSCC, every subblock within any codeword has the same composition, and this composition is chosen to meet the subblock energy constraint (5). The capacity using CSCC (given by (11)) is achieved by choosing that subblock-composition in Γ_B^L (given by (10)) which maximizes the information rate. We will see that rates greater than $C_{CSCC}^L(B)$ can be achieved while still meeting the subblock energy constraint (1).

We first review known results when constraints are placed on the entire codeword (with no subblock constraints) [10], [46]. Let $X_1^n = (X_1, X_2, \dots, X_n)$ denote any codeword of length n . If we impose the average energy constraint on codewords,

$$\frac{1}{n} \sum_{i=1}^n b(X_i) \geq B, \quad (53)$$

then the channel capacity with this constraint is [10], [46]

$$\max_{P_X: \mathbb{E}_{P_X}[b(X)] \geq B} I(X; Y). \quad (54)$$

Information rates arbitrarily close to this capacity can be achieved by making the codeword length sufficiently large. Moreover, if P_X^* is an input distribution which maximizes (54), then this capacity can be achieved by a sequence of CCCs with codeword composition tending to P_X^* [45], [46]. Thus, if $C_{CCC}(B)$ denotes the capacity using CCC when the average energy per symbol is constrained to be at least B , then

$$C_{CCC}(B) = \max_{P: \mathbb{E}_P[b(X)] \geq B} C_{CCC}(P) \quad (55)$$

$$= \max_{P_X: \mathbb{E}_{P_X}[b(X)] \geq B} I(X; Y). \quad (56)$$

Thus the capacity with codeword constraints can be achieved by restricting the codewords to have a fixed composition. This is possible because for a given transmission rate, the codebook size increases exponentially with codeword length n while the number of different types of sequences only increase polynomially with n .

We will now show that contrary to the case with codeword constraints, when the constraints are applied to fixed sized subblocks then information rates can, in general, be increased by *not* restricting the subblocks to have a fixed composition. Towards this, we define a *subblock energy-constrained code* (SECC) as a code which satisfies the subblock energy constraint given by (1). Since all subblocks in SECC satisfy (1), the composition of each subblock belongs to the set Γ_B^L .

Let $C_{SECC}^L(B)$ denote the capacity using SECC with subblock length L and average energy per symbol at least B . Similar to CCCC, the L uses of the channel in case of SECC induce a vector channel with input alphabet

$$\mathcal{A} = \bigcup_{P \in \Gamma_B^L} \mathcal{T}_P^L, \quad (57)$$

output alphabet \mathcal{Y}^L , and channel transition probabilities given by (6). Since each subblock may be chosen independently,

$$C_{SECC}^L(B) = \max_{X_1^L \in \mathcal{A}} \frac{I(X_1^L; Y_1^L)}{L}, \quad (58)$$

where the maximization is over the probability distribution of input vectors in \mathcal{A} . For a *noiseless* q -ary channel ($\mathcal{X} = \mathcal{Y} = \{0, 1, \dots, q-1\}$, $W(i|i) = 1, i \in \mathcal{X}$), it is easy to check that SECC capacity is achieved by the uniform distribution of X_1^L over \mathcal{A} . Thus for the noiseless channel, we have $C_{SECC}^L(B) = \log |\mathcal{A}|/L$.

For CCCC, the induced vector channel was symmetric (irrespective of the underlying (scalar) DMC being symmetric or not), and hence the capacity was achieved with a uniform distribution over the input alphabet. In contrast, in case of SECC the induced vector channel need not be symmetric even when the underlying DMC is symmetric. This is formalized in the following theorem which is proved by providing a counterexample.

Theorem 8. *Uniform distribution of X_1^L over \mathcal{A} may not achieve SECC capacity even when the underlying DMC is symmetric.*

Proof: See Appendix H. ■

Finding the probability distribution which achieves the maximum in (58) is not straightforward, in general. If $U_{\mathcal{A}}$ denotes the uniform distribution of X_1^L over \mathcal{A} , then the maximum information rate achievable with $U_{\mathcal{A}}$, denoted $C_{U_{\mathcal{A}}}^L(B)$, acts as a lower bound for $C_{SECC}^L(B)$. Since a CCCC can be viewed as a SECC where the input vectors have the same composition, it follows that $C_{CCCC}^L(B)$ is also a lower bound for $C_{SECC}^L(B)$. Thus we have

$$C_{SECC}^L(B) \geq \max\{C_{CCCC}^L(B), C_{U_{\mathcal{A}}}^L(B)\}. \quad (59)$$

The following proposition is useful in reducing the computational complexity of $C_{U_{\mathcal{A}}}^L(B)$.

Proposition 3. *For a random input vector X_1^L uniformly distributed over \mathcal{A} with corresponding output vector Y_1^L , the pairwise joint probability, for $1 \leq i \leq L$, satisfies*

$$P_{XY}(X_i = x, Y_i = y) = \sum_{P \in \Gamma_B^L} \frac{|\mathcal{T}_P^L|}{|\mathcal{A}|} P(x) W(y|x) \quad (60)$$

Proof: When X_1^L is uniformly distributed over \mathcal{A} ,

$$\Pr(X_1^L \in \mathcal{T}_P^L) = \frac{|\mathcal{T}_P^L|}{|\mathcal{A}|}. \quad (61)$$

From Prop. 1 it follows that

$$\Pr(X_i = x, Y_i = y \mid X_1^L \in \mathcal{T}_P^L) = P(x)W(y|x). \quad (62)$$

Finally (60) follows from (61) and (62) since $P_{XY}(X_i = x, Y_i = y)$ is equal to

$$\sum_{P \in \Gamma_B^L} \Pr(X_1^L \in \mathcal{T}_P^L) \Pr(X_i = x, Y_i = y \mid X_1^L \in \mathcal{T}_P^L). \quad (63)$$

■

Another useful observation with SECC is that if y_1^L and \tilde{y}_1^L are two output vectors having the same composition, then the columns of the induced vector channel transition matrix corresponding to y_1^L and \tilde{y}_1^L are permutations of each other. This follows from arguments similar to those presented in Appendix A for CSCC. Thus, if X_1^L is distributed uniformly over \mathcal{A} , then for y_1^L and \tilde{y}_1^L having the same composition, we have

$$P_{Y_1^L}(\tilde{y}_1^L) = P_{Y_1^L}(y_1^L) = \frac{1}{|\mathcal{A}|} \sum_{x_1^L \in \mathcal{A}} W^L(y_1^L | x_1^L). \quad (64)$$

The next proposition gives a computationally efficient expression for $C_{U_A}^L(B)$.

Proposition 4. $C_{U_A}^L(B)$ can be expressed as

$$\frac{1}{L} \sum_{Q \in \mathcal{Q}_L} |\mathcal{T}_Q^L| P_{Y_1^L}(y_1^L) \log \frac{1}{P_{Y_1^L}(y_1^L)} - H(Y|X), \quad (65)$$

where \mathcal{Q}_L is the set of all compositions for output vectors of length L , only one representative output vector y_1^L is chosen from every type class \mathcal{T}_Q^L , $P_{Y_1^L}(y_1^L)$ is given by (64), and $H(Y|X)$ is evaluated using the joint pairwise probability distribution given by (60).

Proof: For a DMC, we have

$$C_{U_A}^L(B) = \frac{1}{L} \left(H(Y_1^L) - \sum_{i=1}^L H(Y_i | X_i) \right), \quad (66)$$

where the probability of $y_1^L \in \mathcal{Y}^L$ is given by (64). Thus, (65) follows from (66), (60) and the observation that output vectors with the same composition have equal probability when input subblocks are uniformly distributed over \mathcal{A} . ■

As discussed earlier, the energy requirement per subblock is stricter than the average energy requirement per codeword. Hence, the capacity using codes with subblock-constraint (1) is less

than the capacity using codes with codeword constraint (53). Since CCCs achieve capacity with codeword constraint [46], we have

$$C_{SECC}^L(B) \leq C_{CCC}(B). \quad (67)$$

From (59) and (67) it follows that $C_{SCC}^L(B) \leq C_{SECC}^L(B) \leq C_{CCC}(B)$. Further, using (34) it follows that SECC capacity tends to CCC capacity as $L \rightarrow \infty$. We will compare these capacities for different cases in the numerical results section.

VI. REAL-TIME INFORMATION TRANSFER

So far, we could ensure real-time energy transfer to the receiver by placing constraints on the subblock-composition. For information transfer, although joint decoding of all the subblocks within a codeword is preferred for reducing the probability of error, it also causes delay in information arrival.

For enabling real-time information transfer, the receiver may decode each subblock independently, and thus avoid waiting for arrival of future subblocks. Here, since the subblock decoding proceeds the instant that subblock has been completely received, the information transfer delay is only due to subblock transfer time and the corresponding decoding delay.

When each subblock within the transmitted sequence is decoded independent of other subblocks, then each subblock may itself be viewed as a codeword. We will refer to the independent decoding of subblocks as *local subblock decoding* (LSD). We remark that this subblock based decoding is distinct from decoding for locally decodable codes that allows any bit of the message to be decoded with high probability by only querying a small number of received bits [50].

A. Local Subblock Decoding

In case of local subblock decoding, each subblock may be treated as an independent codeword since every subblock is decoded independently. We are interested in estimating achievable rates with bounded error probability when local subblock decoding is employed. We now provide a short review of an existing result on achievable rates for constant composition finite blocklength codes. This result will then be used (in Sec. VII) to compare rates between local (independent) subblock decoding and joint subblock decoding.

Let $M^*(n, \epsilon)$ denote the maximum size of length- n constant composition code for a DMC with average error probability no larger than ϵ . When the composition of codewords is equal

to an input probability distribution which maximizes the mutual information and the channel satisfies some regularity conditions, then [51]–[53]

$$\log M^*(n, \epsilon) = nC - \sqrt{nV}Q^{-1}(\epsilon) + \frac{1}{2}\log n + O(1) \quad (68)$$

where C is the channel capacity, V is the *information variance*, and Q is the Gaussian Q -function [52]. We remark that V is also termed *channel dispersion* in literature [54]. Early results on finite blocklength capacity for memoryless symmetric channels are due to Weiss [55], which were generalized for the DMC and strengthened by Strassen [56].

When each codeword has equal number of ones and zeros, the achievable rate in bits per channel use for BSC with crossover probability p using CCC is approximated as [51]:

$$\frac{\log_2 M^*(n, \epsilon)}{n} \approx C - \sqrt{\frac{p(1-p)}{n}} \log_2 \frac{1-p}{p} Q^{-1}(\epsilon) + \frac{1}{2n} \log_2 n, \quad (69)$$

with $C = 1 + p \log_2 p + (1-p) \log_2 (1-p)$.

VII. NUMERICAL RESULTS AND DISCUSSION

In this section, we provide examples highlighting the tradeoff between delivery of sufficient energy to the receiver and achieving high information transfer rates. These results are used to draw meaningful insights into choice of subblock length and subblock composition as a function of required energy per symbol at the receiver.

Fig. 3 plots $C_{SCC}^L(B)$ as a function of B for different values of L for a BSC with crossover probability $p_0 = 0.1$. The b -values are assumed to be $b(0) = 0$ and $b(1) = 1$. These b -values reflect the case of on-off keying where bit-1 (bit-0) is represented by the presence (absence) of a carrier signal. Fig. 3 shows that, in general, the value of information rate given by $C_{SCC}^L(B)$ increases with an increase in the subblock length L , for a given B . This is because an increase in L leads to greater choice for input symbols within a subblock. Note that the smaller the value of L , the greater the uniformity in energy distribution within a codeword. The reduction in capacity due to choice of smaller L is the price we pay for providing smoother energy content.

The plot for $L = \infty$ is evaluated using (54); this follows from (11), (33), (55), and the fact that $\lim_{L \rightarrow \infty} r(L, P) = 0$. Thus the curve corresponding to $L = \infty$ is same as the $C_{CCC}(B)$ curve. This curve is a non-increasing concave function of B for $0 \leq B \leq b_{\max}$. This claim can be proved using the approach in [10]. It is non-increasing since the feasibility set Γ_B^L will only become smaller on increasing B . The concavity of $C_{CCC}(B)$ follows from the concavity

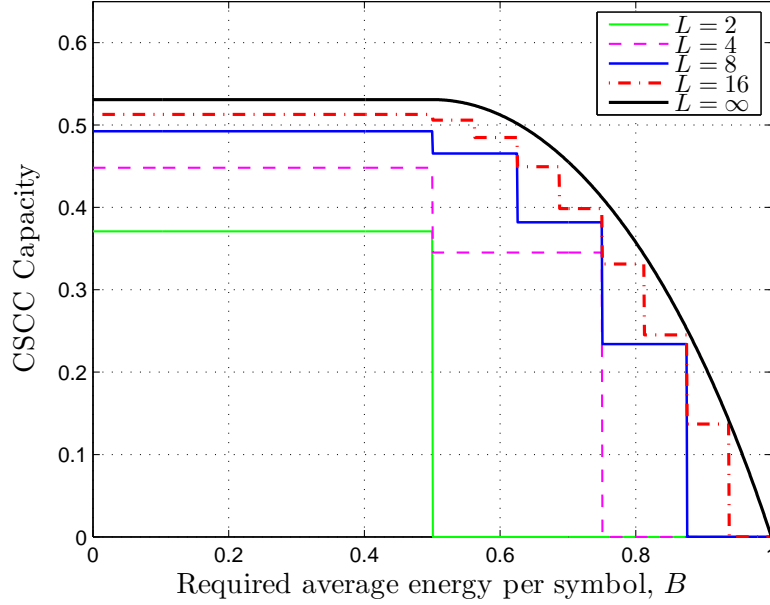


Fig. 3. Plot of $C_{SCC}^L(B)$ versus B for BSC with crossover probability $p_0 = 0.1$, $b(0) = 0$, $b(1) = 1$.

of $I(X; Y)$ as a function of probability distribution of X and the fact that for $0 < \alpha < 1$, the conditions $\mathbb{E}_{P_1}[b(X)] \geq B_1$ and $\mathbb{E}_{P_2}[b(X)] \geq B_2$ imply that

$$\mathbb{E}_{\alpha P_1 + (1-\alpha)P_2}[b(X)] \geq \alpha B_1 + (1 - \alpha)B_2. \quad (70)$$

The non-increasing concave nature of the capacity-power function was used in [57] to show the suboptimality of a time-sharing approach to energy and information transfer.

The CSCC capacity is plotted in Fig. 4 for a BSC as a function of the receiver energy buffer size, E_{max} , with $B = 0.5$. The subblock length L is chosen as a function of E_{max} to satisfy (27). Since L increases with increasing values of E_{max} , the CSCC capacity is an increasing function of E_{max} . For $p_0 = 0.1$, the CSCC capacity is limited by the relatively high value of the crossover probability, rather than the subblock length, with capacity remaining almost constant as E_{max} is increased beyond 10. On the other hand, for $p_0 = 0.01$, the CSCC capacity is limited by the subblock length (since ‘noise’ is weak). From (27) we observe that the subblock length tends to infinity as E_{max} tends to infinity, and hence the CSCC capacity corresponding to $E_{max} \rightarrow \infty$ is equal to $C_{CCC}(B)$.

Fig. 5 plots the rate penalty incurred by using CSCC instead of CCC, for a BSC with crossover probability p_0 , $L = 16$, and $\Pr(0) = \Pr(1) = 0.5$. As discussed in Sec. IV-A, the upper bound

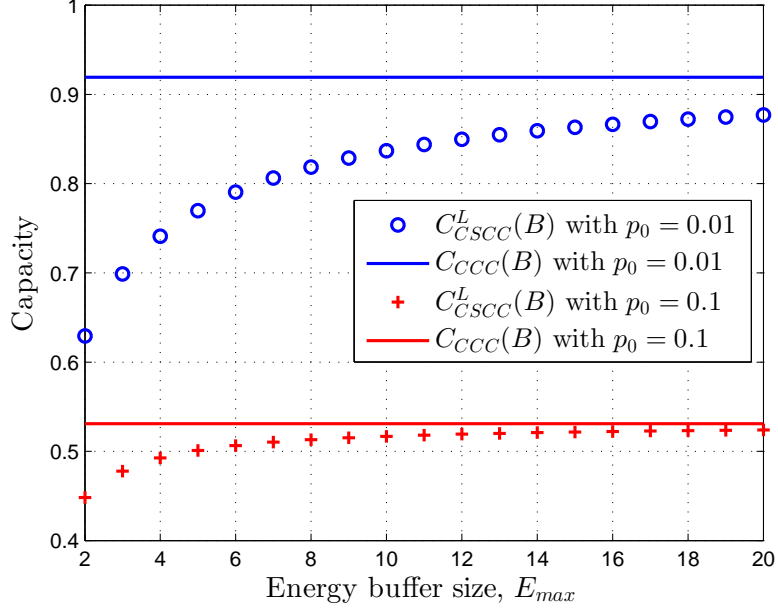


Fig. 4. Plot of CSCC capacity versus receiver energy buffer size, E_{max} , with $B = 0.5, b(0) = 0, b(1) = 1$ for BSC with crossover probability $p_0 = \{0.01, 0.1\}$.

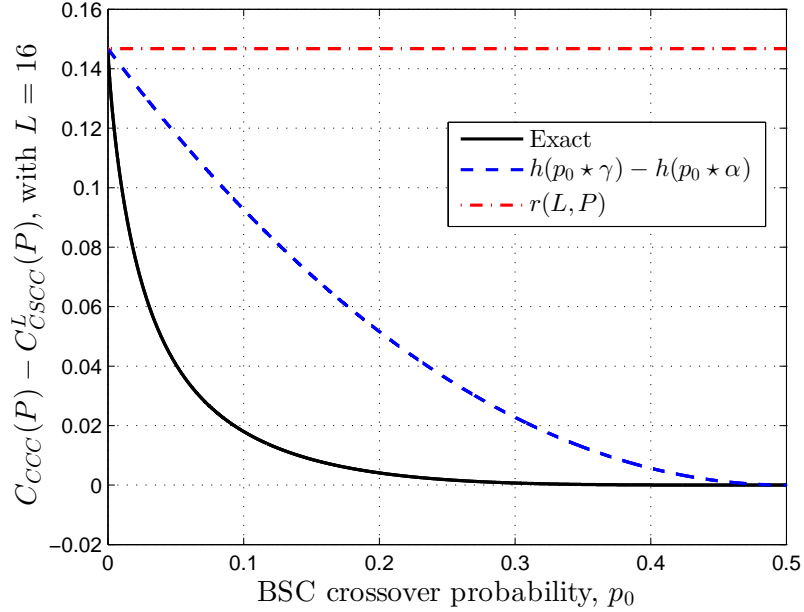


Fig. 5. Plot of $C_{CCC}(P) - C_{CSCC}^L(P)$ as a function of BSC crossover probability p_0 for $L = 16$ and $\Pr(0) = \Pr(1) = 0.5$.

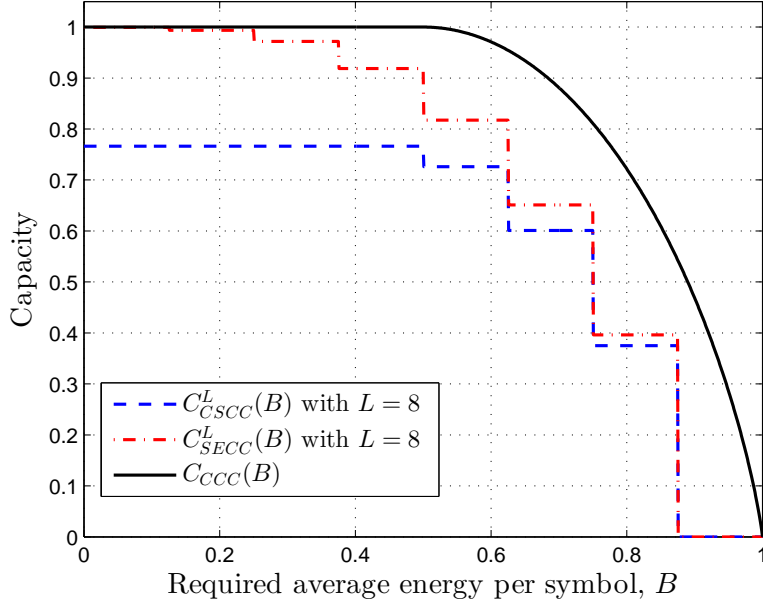


Fig. 6. Comparison of capacity of different schemes for a noiseless binary channel with $b(0) = 0, b(1) = 1$.

on the rate penalty given by $r(L, P)$ is shown to be close to the exact value when $p_0 \approx 0$. Note that $r(L, P)$ is independent of the underlying channel. A tighter bound on the rate penalty given by $h(p_0 \star \gamma) - h(p_0 \star \alpha)$ is also plotted (see Theorem 4). These bounds are useful in estimating the rate penalty for large values of L when the computational complexity of $C_{CSCC}^L(P)$ becomes high. The bounds on rate penalty may also be used to bound the exact value of $C_{CSCC}^L(P)$ for large L .

Fig. 6 compares the capacity of CSCC and SECC for a noiseless binary channel with $b(0) = 0, b(1) = 1$ and subblock length $L = 8$. Note that the capacity curve for CCC may be viewed as the CSCC capacity curve corresponding to $L = \infty$. Fig. 6 highlights the potential of improving the CSCC capacity by using SECCs and allowing different subblocks to have different compositions while still meeting the subblock energy constraint (1). With SECCs, the capacity for a noiseless channel is achieved by a uniform distribution of input vectors and can thus be efficiently computed using (65).

Fig. 7 compares capacity of different schemes for $L = 8$ and $B = 0.6$, as a function of BSC crossover probability p_0 . It shows that for $p_0 < 0.05$, the capacity with uniform distribution over the set of length L vectors which satisfy the subblock energy constraint (1), is higher

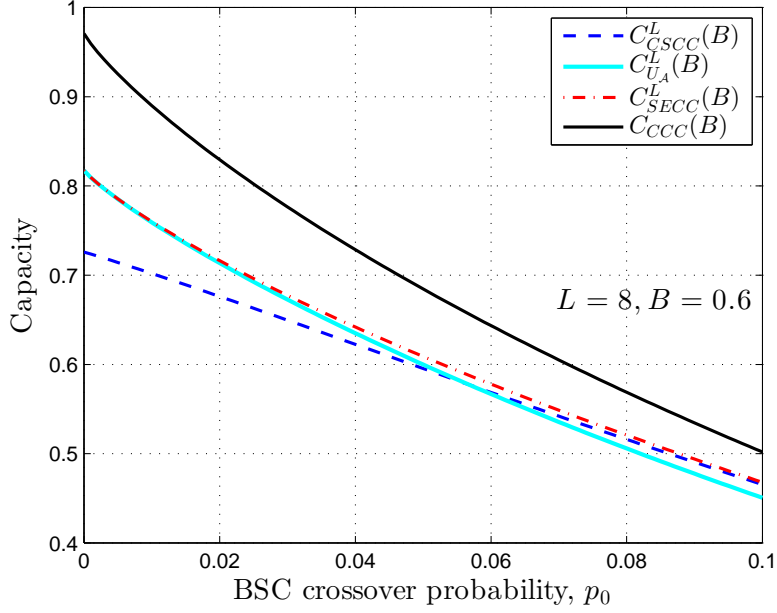


Fig. 7. Comparison of capacity of different schemes for $L = 8, B = 0.6$, as a function of BSC crossover probability p_0 and $b(0) = 0, b(1) = 1$.

compared to CCCC capacity. However, $C_{U_A}^L(B) < C_{SCC}^L(B)$ for relative higher values of p_0 . This observation emphasizes the fact that merely adding more types is not sufficient to increase capacity compared to CCCC; we need to choose an appropriate distribution over the enlarged alphabet as well. In Fig. 7, we used the Blahut-Arimoto algorithm [58], [59] to compute the exact SECC capacity, $C_{SECC}^L(B)$.

Fig. 8 compares achievable rates using local subblock decoding (LSD) with rates using joint subblock decoding for a BSC with crossover probability $p_0 = 0.11$ when each subblock has equal number of zeros and ones (that is, $P(0) = P(1) = 0.5$). In case of CCCC with LSD, each subblock may itself be viewed as a codeword and so the achievable rate is approximated by (69) with $n = L$. The achievable rates with LSD are obtained using (69) and seen to fall significantly as the desired probability of error, ϵ , tends to zero. The red curve plots lower bound on $C_{SCC}^L(P)$ obtained using (37). Note that $C_{SCC}^L(P)$ represents the rate with joint subblock decoding for which the probability of error can be brought arbitrarily close to zero by increasing the number of subblocks in a codeword and then jointly decoding the subblocks.

Notice that the rate loss decreases as $\sqrt{1/L}$ with LSD whereas the rate loss with joint decoding decreases as $\log(L)/L$. Ensuring the ability to use energy in real-time imposes less of a penalty

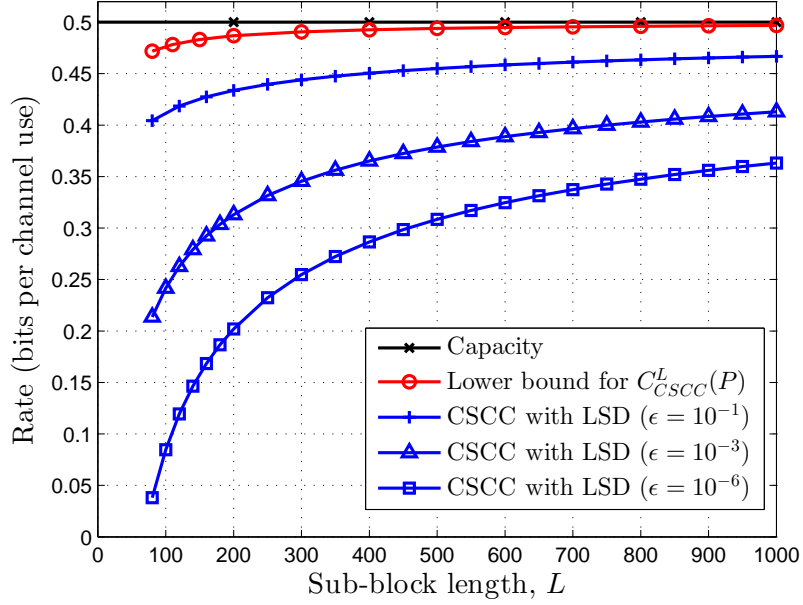


Fig. 8. Rates for a BSC with crossover probability $p = 0.11$.

than the ability to use information in real-time.

VIII. REFLECTIONS

We proposed the use of CSCC codes for providing regular energy content in a patterned energy signal which is used for simultaneous transfer of energy and information. The subblock-composition in CSCC was chosen to maximize the rate of information transfer while ensuring that the fraction of input symbols carrying high energy within every subblock duration are sufficiently large. For characterizing the exact CSCC capacity, we employed a super-letter approach (with each subblock being viewed as a single super-letter in an induced vector-channel) and showed that CSCC capacity computational complexity can be alleviated by exploiting certain symmetry properties.

The super-letter approach can also be applied to compute CSCC error exponent. However, the size of the super-alphabet grows exponentially with subblock length, L , and the cost for computing exact CSCC capacity and error exponent may become prohibitive for large L . In this scenario, the CSCC capacity and error exponent can be estimated by using their respective bounds, derived in Sec. IV, in terms of the capacity and error exponent for constant composition codes. Compared to CCC, the use of CSCCs incurs a rate loss due to the constraint restricting

the subblocks to have the same composition. We showed that the CCCC error exponent is related to the CCC error exponent by the same rate loss term.

We also showed that CCCC capacity can be increased by allowing different subblocks to have different compositions, while still meeting the subblock energy constraint. Although CCCC capacity is shown to be achieved by uniform distribution of super-letters, one may have to resort to numerical techniques (such as the Blahut-Arimoto algorithm) for obtaining a capacity achieving input distribution for the case where different subblocks are permitted to have different compositions.

We provided examples highlighting the tradeoff between delivery of sufficient energy to the receiver and achieving high information transfer rates. It was observed that the ability to use energy in real-time imposes less of penalty than the ability to use information in real-time.

We showed that the subblock length in CCCC can be bounded as a function of the receiver energy storage capacity to avoid energy outage at the receiver. In scenarios where the energy harvested at the receiver upon transmission of an input symbol varies over time, it will be appealing to analyze bounds on subblock length which apply energy arrival statistics to ensure that the energy outage probability is lower than a certain threshold. Future work may also be carried on extending CCCC capacity results to other channel models, such as the AWGN channel where the average transmit power is also constrained.

Other than the application of simultaneous energy and information transfer, CCCCCs are also suitable candidates for power line communications due to their ability to provide regular energy content. The CCCC codes may also find application in other diverse fields. For instance, the multiply constant-weight codes (MCWC) proposed in [28] for use in low-cost authentication methods are a special case of CCCC with binary input alphabet. Thus, our capacity results for CCCC can also be employed as a performance benchmark for practical MCWC codes [27].

ACKNOWLEDGMENT

The authors thank Vincent Y. F. Tan for his helpful comments.

APPENDIX A

PROOF OF THEOREM 1

We will prove Theorem 1 by first proving some simple lemmas and employing Gallager's definition of a symmetric channel [60].

If π denotes any permutation on L letters with

$$\pi(x_1^L) = \pi(x_1, x_2, \dots, x_L) \triangleq (x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(L)}), \quad (71)$$

then we have the following lemmas.

Lemma 3.

$$W^L(\pi(y_1^L)|\pi(x_1^L)) = W^L(y_1^L|x_1^L) \quad (72)$$

Proof: For a DMC, we have

$$\begin{aligned} W^L(\pi(y_1^L)|\pi(x_1^L)) &= \prod_{i=1}^L W(y_{\pi(i)}|x_{\pi(i)}) \\ &= \prod_{i=1}^L W(y_i|x_i) = W^L(y_1^L|x_1^L) \end{aligned}$$

■

Lemma 4. *The following sets are equal*

$$\{\pi(x_1^L)|x_1^L \in \mathcal{T}_P^L\} = \mathcal{T}_P^L \quad (73)$$

Proof: A permutation preserves the composition of a sequence. Thus, π may be viewed as a map $\pi : \mathcal{T}_P^L \rightarrow \mathcal{T}_P^L$. This map is injective by definition of a permutation. Since the set \mathcal{T}_P^L is finite, this map is also surjective and hence (73) follows. ■

Lemma 5. *The following sets are equal*

$$\{W^L(\pi(y_1^L)|x_1^L) : x_1^L \in \mathcal{T}_P^L\} = \{W^L(y_1^L|x_1^L) : x_1^L \in \mathcal{T}_P^L\} \quad (74)$$

Proof: From Lemma 3 we have $W^L(\pi(y_1^L)|x_1^L) = W^L(y_1^L|\pi^{-1}(x_1^L))$. Now (74) follows from Lemma 4. ■

Let the composition of the output vector $y_1^L \in \mathcal{Y}^L$ be Q and let \mathcal{T}_Q^L be the set of all output vectors of length L having composition Q .

Lemma 6. *The following sets are equal*

$$\{W^L(y_1^L|\pi(x_1^L)) : y_1^L \in \mathcal{T}_Q^L\} = \{W^L(y_1^L|x_1^L) : y_1^L \in \mathcal{T}_Q^L\} \quad (75)$$

Proof: Similar to Lemma 5. ■

We recall Gallager's definition [60] of a *symmetric DMC*.

Definition 1. *A DMC is symmetric if the set of outputs can be partitioned into subsets in such a way that for each subset the matrix of transition probabilities (using inputs as rows and outputs of the subsets as columns) has the property that each row is a permutation of each other row and each column (if more than 1) is a permutation of each other column.*

We will show that when CSCC is employed on a DMC, the induced vector-channel is symmetric. Note that the underlying (scalar) channel can be any arbitrary DMC (not necessarily symmetric).

Lemma 7. *When CSCC with subblock length L is employed on any DMC, the induced vector-channel (obtained from L uses of the DMC) is symmetric.*

Proof: The lemma will be proved if we can partition the outputs into subsets such that for each subset the matrix of transition probabilities has the property that each row (column) is a permutation of each other row (column).

We now show that if we partition the outputs into subsets such that each subset contains all the outputs of a given composition, then the symmetry conditions will be satisfied.

If $y_1^L \in \mathcal{T}_Q^L$ and $\tilde{y}_1^L \in \mathcal{T}_Q^L$ for a given composition Q , then since y_1^L and \tilde{y}_1^L have the same composition, we have $\tilde{y}_1^L = \pi(y_1^L)$ for some permutation π . Let \mathcal{T}_P^L be the input alphabet for the induced vector channel using CSCC with subblock-composition P . Then using Lemma 5, we note that the columns of the vector-channel transition matrix corresponding to output subset \mathcal{T}_Q^L are permutations of each other. Similarly, using Lemma 6 we can prove that the corresponding rows are permutations of each other. ■

Theorem 9 ([60, p. 94]). *For a symmetric discrete memoryless channel, capacity is achieved by using the inputs with equal probability.*

Finally, Theorem 1 follows directly from Lemma 7 and Theorem 9. ■

APPENDIX B

PROOF OF THEOREM 2

When L satisfies (27), then $E_{max} \geq 2G$. Since $E(1) \geq G$, the energy level at the start of *every* subblock is at least G (by recursive application of Lemma 1(c)) and sufficiency follows from Lemma 1(b).

Now let $L_1 = \sum_{x \in \mathcal{X}_\triangleleft} LP(x)$, and define

$$P_1(x) = \begin{cases} \frac{P(x)}{\sum_{x \in \mathcal{X}_\triangleleft} P(x)}, & \text{if } x \in \mathcal{X}_\triangleleft \\ 0, & \text{if } x \in \mathcal{X}_\triangleright \end{cases} \quad (76)$$

$$P_2(x) = \begin{cases} 0, & \text{if } x \in \mathcal{X}_\triangleleft \\ \frac{P(x)}{\sum_{x \in \mathcal{X}_\triangleright} P(x)}, & \text{if } x \in \mathcal{X}_\triangleright \end{cases} \quad (77)$$

$$S_1 = \{x_1^L \mid x_1^{L_1} \in \mathcal{T}_{P_1}^{L_1}, x_{L_1+1}^L \in \mathcal{T}_{P_2}^{L-L_1}\} \quad (78)$$

$$S_2 = \{x_1^L \mid x_1^{L-L_1} \in \mathcal{T}_{P_2}^{L-L_1}, x_{L-L_1+1}^L \in \mathcal{T}_{P_1}^{L_1}\}. \quad (79)$$

Clearly $S_1 \subset \mathcal{T}_P^L$, $S_2 \subset \mathcal{T}_P^L$, where S_1 (resp. S_2) denotes the set of subblocks of length L with first (resp. last) L_1 input symbols belonging to $\mathcal{X}_\triangleleft$. Note that $E(1) \geq G$ is necessary to avoid outage because if $E(1) < G$, then outage results when the first subblock in a codeword belongs to S_1 . To prove that (27) is necessary, we will show that when

$$L > \frac{E_{max}}{\sum_{x \in \mathcal{X}_\triangleleft} 2P(x) (B - b(x))}, \quad (80)$$

then CSCC codewords exist which will result in energy outage at the receiver. Here we have

$$G = \sum_{x \in \mathcal{X}_\triangleleft} LP(x) (B - b(x)) > \frac{E_{max}}{2}. \quad (81)$$

Let the first subblock in a given codeword belong to S_2 . Since the last L_1 symbols (within the first subblock) belong to $\mathcal{X}_\triangleleft$, we have $E(L+1) = |E(L-L_1+1) - G|^+$. If there is no outage during the reception of the first subblock,

$$E(L+1) = E(L-L_1+1) - G \leq E_{max} - G < E_{max}/2, \quad (82)$$

where the last inequality follows from (81). Now let the second subblock belong to S_1 . There is no energy outage during the reception of first L_1 symbols within the second subblock if and only if $E(L+1) \geq G$. However, from (82) and (81) it follows that $E(L+1) < E_{max}/2 < G$, and hence outage cannot be avoided in the second subblock. In general, outage results if L satisfies (80), and any two adjacent subblocks in a codeword belongs to S_2 and S_1 , respectively. ■

APPENDIX C

PROOF OF THEOREM 4

The strict inequality $0 < C_{CCC}(P) - C_{CSCC}^L(P)$ follows for BSC with crossover probability $0 < p_0 < 0.5$ because

$$C_{CSCC}^L(P) = \frac{1}{L} [H(Y_1^L) - H(Y_1^L|X_1^L)] \quad (83)$$

$$= \frac{1}{L} \left[\sum_{i=1}^L H(Y_i|Y_1^{i-1}) - \sum_{i=1}^L H(Y_i|X_i) \right] \quad (84)$$

$$\stackrel{(a)}{<} \frac{1}{L} \left[\sum_{i=1}^L H(Y_i) - \sum_{i=1}^L H(Y_i|X_i) \right] \quad (85)$$

$$= C_{CCC}(P), \quad (86)$$

where $Y_1^{i-1} = Y_1 \dots Y_{i-1}$, the strict inequality (a) follows since Y_i is related to Y_1^{i-1} via X_1^{i-1} and X_i . The last equality above follows from Prop. 1 and (28).

For subblock-composition P with $0 < \gamma = \min(P(0), P(1)) \leq 0.5$, the output entropy on a BSC is $H(Y) = h(p_0 \star \gamma)$ and hence

$$C_{CCC}(P) = h(p_0 \star \gamma) - h(p_0). \quad (87)$$

For CSCC, from (29) and definition of α , it follows that

$$\frac{1}{L} H(X_1^L) = H(P) - r(L, P) \quad (88)$$

$$= h(\gamma) - r(L, P) = h(\alpha). \quad (89)$$

Now using (89) and applying Mrs. Gerber's Lemma [47],

$$\frac{1}{L} H(Y_1^L) \geq h(p_0 \star \alpha), \quad (90)$$

and hence

$$C_{CSCC}^L(P) = \frac{1}{L} \left[H(Y_1^L) - \sum_{i=1}^L H(Y_i|X_i) \right] \quad (91)$$

$$\geq h(p_0 \star \alpha) - h(p_0). \quad (92)$$

Using (87) and (92) we have

$$C_{CCC}(P) - C_{CSCC}^L(P) \leq h(p_0 \star \gamma) - h(p_0 \star \alpha) \quad (93)$$

We only have to show that $h(p_0 \star \gamma) - h(p_0 \star \alpha) < r(L, P)$ for completing the proof. Towards this we first observe that when $0 < x \leq 0.5$ and $0 < p_0 < 0.5$, then $p_0 \star x \geq x$. Next we note that the derivative of $h(x)$ satisfies

$$h'(x) = \log \frac{1-x}{x}, \quad (94)$$

and hence $h'(x)$ is a monotonically decreasing function of x for $0 < x \leq 0.5$.

Since $h(\alpha) = h(\gamma) - r(L, P)$, we have

$$\begin{aligned} h(p_0 \star \gamma) - h(p_0 \star \alpha) &< r(L, P) \iff \\ h(p_0 \star \gamma) - h(\gamma) &< h(p_0 \star \alpha) - h(\alpha). \end{aligned} \quad (95)$$

If we define $f(x) = h(p_0 \star x) - h(x)$ for $0 \leq x \leq 0.5$, then we have

$$f'(x) = (1 - 2p_0)h'(p_0 \star x) - h'(x). \quad (96)$$

Hence $f'(x) < 0$ for $0 < x \leq 0.5$ since $h'(x)$ is monotonically decreasing in x and $p_0 \star x \geq x$. This in turn implies that $f(x)$ is a strictly monotonically decreasing function of x . It follows that $f(\gamma) < f(\alpha)$ (since $\alpha < \gamma$) and (95) is satisfied. \blacksquare

APPENDIX D

PROOF OF THEOREM 5

For a BEC with erasure probability ϵ , and $\gamma = P(0)$,

$$C_{CCC}(P) = (1 - \epsilon)h(\gamma). \quad (97)$$

If α is chosen such that $h(\alpha) = h(\gamma) - r(L, P)$, then from (29) it follows that $H(X_1^L)/L = h(\alpha)$. Now applying an extension of MGL for binary input symmetric channels [48], we get $H(Y_1^L)/L \geq (1 - \epsilon)h(\alpha) + h(\epsilon)$. Thus,

$$C_{SCC}^L(P) = \frac{1}{L} \left[H(Y_1^L) - \sum_{i=1}^L H(Y_i|X_i) \right] \geq (1 - \epsilon)h(\alpha), \quad (98)$$

and (39) follows from (97), (98), and definition of α .

APPENDIX E
PROOF OF THEOREM 6

For a Z -channel with $\gamma = \Pr(X = 1)$, $p_0 = \Pr(1 \rightarrow 0)$,

$$C_{CCC}(P) = h(\gamma(1 - p_0)) - \gamma h(p_0). \quad (99)$$

If $0 \leq \alpha \leq 0.5$ is chosen such that $h(\alpha) = h(\gamma) - r(L, P)$, then from (29) it follows that $H(X_1^L)/L = h(\alpha)$. Now applying the extension of MGL for memoryless asymmetric binary-input, binary-output channels [49], we get $H(Y_1^L)/L \geq h(\alpha(1 - p_0))$. Thus,

$$C_{SCC}^L(P) = h(\alpha(1 - p_0)) - \gamma h(p_0), \quad (100)$$

and (40) follows from (99) and (100).

APPENDIX F
PROOF OF LEMMA 2

We first note that the functions $D(V||W|P)$ and $I(P, V)$ are convex functions of V , while the constraint $\sum_{y \in \mathcal{Y}} V(y|x) = 1$ for $x \in \mathcal{X}$ is linear in V . Thus the problem of minimization of $D(V||W|P)$ over V subject to $I(P, V) \leq R$ and $\sum_{y \in \mathcal{Y}} V(y|x) = 1$ is a convex optimization problem and can be solved by the method of Lagrange multipliers [61].

Secondly, note that the functions $D(V||W|P)$ and $I(P, V)$ depend on V only through those $V(y|x)$ for which $P(x) > 0$. Thus we assume, without loss of generality, that $P(x) > 0, \forall x \in \mathcal{X}$.

Now consider the Lagrangian $\xi(V)$ given by

$$\begin{aligned} \xi(V) = & D(V||W|P) + \lambda(I(P, V) - R) \\ & + \sum_x \nu_x \left(\sum_y V(y|x) - 1 \right) \end{aligned} \quad (101)$$

where $\lambda \geq 0$. On setting the partial derivative of $\xi(V)$ with respect to $V(y|x)$ equal to zero, we get

$$\begin{aligned} 0 = & P(x) \log \frac{V(y|x)}{W(y|x)} + P(x) \log e \\ & + \lambda P(x) \log \frac{V(y|x)}{PV(y)} + \nu_x. \end{aligned} \quad (102)$$

On substituting $s = \lambda/(1 + \lambda)$, the above equation can be equivalently be expressed as

$$V(y|x) = W(y|x)^{1-s} PV(y)^s \exp\left(-\frac{(1-s)(P(x) \log e + \nu_x)}{P(x)}\right) \quad (103)$$

Since $\sum_y V(y|x) = 1$, using (103) we get

$$\exp\left(\frac{(1-s)(P(x)\log e + \nu_x)}{P(x)}\right) = \sum_y W(y|x)^{1-s} PV(y)^s, \quad (104)$$

and finally using (103), (104), we have

$$V(y|x) = \frac{W(y|x)^{1-s} PV(y)^s}{\sum_y W(y|x)^{1-s} PV(y)^s}. \quad (105)$$

Note that since $s = \lambda/(1 + \lambda)$, the value of s ranges from 0 to 1 as λ varies from 0 to ∞ . By the complementary slackness property [61], we have

$$\lambda(I(P, V) - R) = 0. \quad (106)$$

Since $\lambda = 0$ implies $s = 0$ and hence $V = W$ (using (105)), it follows that $I(P, W) = I(P, V) \leq R$. This contradicts the assumption in Lemma 2 that $R < I(P, W)$ and hence λ is strictly greater than zero. The proof is complete by noting that conditions $\lambda > 0$ and (106) imply $I(P, V) = R$. \blacksquare

APPENDIX G

PROOF OF THEOREM 7

The M messages to be transmitted are assumed equiprobable. All input sequences of length n with constant subblock-composition P are assigned equal probabilities and the i th message is mapped to a randomly selected input sequence for $1 \leq i \leq M$. The decoder knows the mapping used by the encoder and uses maximum likelihood (ML) decoding.

The proof uses Fano's approach [45] to upper bound the probability of error by employing *tilted distribution* which is summarized next.

Let A be a discrete ensemble consisting of points a_1, \dots, a_m with probability distribution $\tilde{P}(a)$. If ϕ denotes a random variable associated with this ensemble, and $\gamma(s) := \log \sum_A \exp(s\phi(a)) \tilde{P}(a)$, then a family of *tilted distributions* are:

$$Q(a) := \exp(s\phi(a) - \gamma(s)) \tilde{P}(a). \quad (107)$$

Note that for a fixed s , the derivative $\gamma'(s)$ (resp. $\gamma''(s)$) denotes the mean (resp. variance) of the random variable ϕ with respect to the tilted probability distribution Q . For $s = 0$, we have $Q = P$, and thus $\gamma'(0)$ (resp. $\gamma''(0)$) denotes the true mean (resp. variance) of ϕ .

Let $n_i, i = 1, \dots, K$ be positive integers and $n = \sum_{i=1}^K n_i$. Define the subsets $S_1 = \{1, \dots, n_1\}$ and $S_k = \{l \in \mathbb{N} \mid \sum_{i=1}^{k-1} n_i < l \leq \sum_{i=1}^k n_i\}$ for $2 \leq k \leq K$. Let $\alpha^n := \alpha_1 \cdots \alpha_n$ denote a

sequence of n independent events. For any $i \in S_k, k = 1, \dots, K$, let $\phi_k(\alpha_i)$ be the random variable associated with the event α_i and $\tilde{P}_k(a)$ be the probability $\Pr[\alpha_i = a]$. For $k = 1, \dots, K$, define

$$\gamma_k(s) := \log \sum_A \exp(s\phi_k(a)) \tilde{P}_k(a) \quad (108)$$

$$\gamma(s) := \sum_{k=1}^K \frac{n_k}{n} \gamma_k(s) \quad (109)$$

$$Q_k(a) := \exp(s\phi_k(a) - \gamma_k(s)) \tilde{P}_k(a) \quad (110)$$

Define the sum of random variables,

$$\Phi(\alpha^n) := \sum_{k=1}^K \sum_{i \in S_k} \phi_k(\alpha_i). \quad (111)$$

whose tail probability is given by the following lemma.

Lemma 8. (*[45, p. 265]*) Assume $\gamma(s)$ and its first and second derivative are finite in the interval $s_1 < s < s_2$ including $s = 0$. If t is a real number with $\gamma'(s_1) < t < \gamma'(s_2)$, then the tail probabilities of $\Phi(\alpha_n)$ satisfy the following inequalities:

$$\Pr[\Phi(\alpha^n) \leq nt] \leq \exp(-n\beta), \quad \gamma'(s_1) < t < \gamma'(0) = \bar{\phi} \quad (112)$$

$$\Pr[\Phi(\alpha^n) > nt] \leq \exp(-n\beta), \quad \gamma'(0) = \bar{\phi} \leq t < \gamma'(s_2) \quad (113)$$

where

$$\beta = s\gamma'(s) - \gamma(s) = \sum_{k=1}^K \frac{n_k}{n} \sum_A Q_k(a) \log \frac{Q_k(a)}{\tilde{P}_k(a)} \geq 0 \quad (114)$$

with s chosen such that $\gamma'(s) = t$.

Define the distance between $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ as

$$\tilde{D}(x, y) = \log \frac{f(y)}{W(y|x)}, \quad (115)$$

where $f(y)$ is a positive function of y with $\sum_{\mathcal{Y}} f(y) = 1$. Similarly, the distance between two sequences \mathbf{u} and \mathbf{v} is

$$\tilde{D}(\mathbf{u}, \mathbf{v}) = \sum_{\mathcal{X}, \mathcal{Y}} n(x, y) \tilde{D}(x, y) = \log \frac{F(\mathbf{v})}{W^n(\mathbf{v}|\mathbf{u})} \quad (116)$$

where $n(x, y)$ is the number of letter pairs (x, y) in (\mathbf{u}, \mathbf{v}) ,

$$F(\mathbf{v}) = \prod_{\mathcal{Y}} f(y)^{n(y)}, \quad W^n(\mathbf{v}|\mathbf{u}) = \prod_{\mathcal{X}, \mathcal{Y}} W(y|x)^{n(x, y)}. \quad (117)$$

If \mathbf{u} is the input sequence and \mathbf{v} is the corresponding output sequence, then an error may occur with ML decoding when one of the other $M - 1$ messages is represented by another input sequence \mathbf{u}' for which $W^n(\mathbf{v}|\mathbf{u}') \geq W^n(\mathbf{v}|\mathbf{u})$, or equivalently $\tilde{D}(\mathbf{u}', \mathbf{v}) \leq \tilde{D}(\mathbf{u}, \mathbf{v})$. The following lemma gives an upper bound on the probability of error.

Lemma 9. (*[45, p. 307]*) *Let D_0 be an arbitrary constant, \mathbf{u}_0 be a particular transmitted codeword, \mathbf{u} be another randomly chosen input sequence, and \mathbf{v} be an output sequence. The average probability of error satisfies the inequality $P_e < MP_1 + P_2$, where:*

$$P_1 = \Pr[\tilde{D}(\mathbf{u}_0, \mathbf{v}) \leq D_0, \tilde{D}(\mathbf{u}, \mathbf{v}) \leq \tilde{D}(\mathbf{u}_0, \mathbf{v})], \quad (118)$$

$$P_2 = \Pr[\tilde{D}(\mathbf{u}_0, \mathbf{v}) > D_0]. \quad (119)$$

For CCC, P_e is independent of \mathbf{u}_0 .

Since a CSCC is also a CCC, Lemma 9 will be used to bound the error probability for CSCC, while Lemma 8 will be used to compute P_2 (119).

We now define some terms and notation which will be used later. We say that sequences \mathbf{v} and \mathbf{v}' , each comprising of m subblocks of length L , have the *same subblock-composition* if the i th subblock in sequences \mathbf{v} and \mathbf{v}' has the same composition, for $1 \leq i \leq m$. The composition of the i th subblock of a sequence pair (\mathbf{u}, \mathbf{v}) is defined as a matrix whose (j, k) entry is equal to $n_i(x_j, y_k)/L$ where $n_i(x_j, y_k)$ is the number of letter pairs (x_j, y_k) in the i th subblock of (\mathbf{u}, \mathbf{v}) . The *subblock-composition of a sequence pair* (\mathbf{u}, \mathbf{v}) is defined as a length m vector whose i th entry is the composition of the i th subblock of (\mathbf{u}, \mathbf{v}) .

The following lemma compares distances between sequences having the same subblock-composition. This lemma will be used to bound P_1 (118).

Lemma 10. *Let \mathbf{u} and \mathbf{v} be two particular sequences with elements in \mathcal{X} and \mathcal{Y} , respectively. Select equiprobably at random a sequence \mathbf{u}' having the same subblock-composition as \mathbf{u} , and a sequence \mathbf{v}' having the same subblock-composition as \mathbf{v} . Then*

$$\Pr[\tilde{D}(\mathbf{u}', \mathbf{v}) \leq \tilde{D}(\mathbf{u}, \mathbf{v})] = \Pr[\tilde{D}(\mathbf{u}, \mathbf{v}') \leq \tilde{D}(\mathbf{u}, \mathbf{v})] \quad (120)$$

Proof: Let $n_i(x)$ (resp. $n_i(y)$) denote the number of occurrences of x (resp. y) in the i th subblock of \mathbf{u} (resp. \mathbf{v}). Let $T_{\mathbf{u}, \mathbf{v}}$ be the set of distinct subblock-compositions for sequence pairs

$(\mathbf{u}', \mathbf{v}')$ for which $\tilde{D}(\mathbf{u}', \mathbf{v}') \leq \tilde{D}(\mathbf{u}, \mathbf{v})$ and \mathbf{u}' (resp. \mathbf{v}') has the same subblock-composition as \mathbf{u} (resp. \mathbf{v}).

The number of sequences \mathbf{u}' , having the same subblock-composition as \mathbf{u} , for which $\tilde{D}(\mathbf{u}', \mathbf{v}) \leq \tilde{D}(\mathbf{u}, \mathbf{v})$ is

$$\sum_{T_{\mathbf{u}, \mathbf{v}}} \prod_{i=1}^m \left(\frac{\prod_{\mathcal{Y}} n_i(y)!}{\prod_{\mathcal{X}, \mathcal{Y}} n_i(x, y)!} \right) \quad (121)$$

The total number of sequences \mathbf{u}' having the same subblock-composition as \mathbf{u} is

$$\prod_{i=1}^m \left(\frac{L!}{\prod_{\mathcal{X}} n_i(x)!} \right) = \frac{(L!)^m}{\prod_{i=1}^m \prod_{\mathcal{X}} n_i(x)!} \quad (122)$$

Thus, $\Pr[\tilde{D}(\mathbf{u}', \mathbf{v}) \leq \tilde{D}(\mathbf{u}, \mathbf{v})]$ is equal to the ratio of (121) to (122).

The number of sequences \mathbf{v}' , having the same subblock-composition as \mathbf{v} , for which $\tilde{D}(\mathbf{u}, \mathbf{v}') \leq \tilde{D}(\mathbf{u}, \mathbf{v})$ is

$$\sum_{T_{\mathbf{u}, \mathbf{v}}} \prod_{i=1}^m \left(\frac{\prod_{\mathcal{X}} n_i(x)!}{\prod_{\mathcal{X}, \mathcal{Y}} n_i(x, y)!} \right) \quad (123)$$

The total number of sequences \mathbf{v}' having the same subblock-composition as \mathbf{v} is

$$\prod_{i=1}^m \left(\frac{L!}{\prod_{\mathcal{Y}} n_i(y)!} \right) = \frac{(L!)^m}{\prod_{i=1}^m \prod_{\mathcal{Y}} n_i(y)!} \quad (124)$$

Thus, $\Pr[\tilde{D}(\mathbf{u}, \mathbf{v}') \leq \tilde{D}(\mathbf{u}, \mathbf{v})]$ is the ratio of (123) to (124) which is equal to the ratio of (121) to (122). ■

We now proceed with the main steps leading to the proof of Theorem 7.

From Lemma 9 it follows that the probability of error (averaged over mappings from messages to codewords) satisfies $P_e < MP_1 + P_2$, where P_2 can be bounded using Lemma 8 so that

$P_2 < \exp(-n\beta)$, with

$$\beta = s\gamma'(s) - \gamma(s) = D(V||W|P), \quad (125)$$

$$\gamma(s) = \sum_{\mathcal{X}} P(x) \log \sum_y \exp \left(s\tilde{D}(x, y) \right) W(y|x), \quad (126)$$

$$\begin{aligned} V(y|x) &= \frac{\exp \left(s\tilde{D}(x, y) \right) W(y|x)}{\sum_y \exp \left(s\tilde{D}(x, y) \right) W(y|x)}; \quad s \geq 0 \\ &= \frac{W(y|x)^{1-s} f(y)^s}{\sum_y W(y|x)^{1-s} f(y)^s}; \quad s \geq 0, \end{aligned} \quad (127)$$

where $\tilde{D}(x, y)$ is defined in (115) and s is chosen such that

$$\gamma'(s) = \sum_{\mathcal{X}, \mathcal{Y}} P(x) V(y|x) \tilde{D}(x, y) = \frac{D_0}{n}. \quad (128)$$

Next, we derive an upper bound for P_1 . If \mathbf{u}_0 is the transmitted sequence, \mathbf{u}' is a randomly chosen sequence having the same subblock-composition as \mathbf{u}_0 , V_0 is the set of sequences \mathbf{v} for which $\tilde{D}(\mathbf{u}_0, \mathbf{v}) \leq D_0$, $V'_\mathbf{v}$ is the set of sequences \mathbf{v}' that have the same subblock-composition as \mathbf{v} , $V'_{0\mathbf{v}}$ is the subset of $V'_\mathbf{v}$ for which $\tilde{D}(\mathbf{u}_0, \mathbf{v}') \leq \tilde{D}(\mathbf{u}_0, \mathbf{v})$, then:

$$\begin{aligned} P_1 &= \sum_{V_0} W^n(\mathbf{v}|\mathbf{u}_0) Pr[\tilde{D}(\mathbf{u}', \mathbf{v}) \leq \tilde{D}(\mathbf{u}_0, \mathbf{v})] \\ &\stackrel{(a)}{=} \sum_{V_0} W^n(\mathbf{v}|\mathbf{u}_0) Pr[\tilde{D}(\mathbf{u}_0, \mathbf{v}') \leq \tilde{D}(\mathbf{u}_0, \mathbf{v})] \\ &= \sum_{V_0} W^n(\mathbf{v}|\mathbf{u}_0) \frac{|V'_{0\mathbf{v}}|}{|V'_\mathbf{v}|} \\ &\stackrel{(b)}{=} \sum_{V_0} \frac{\sum_{V'_{0\mathbf{v}}} W^n(\mathbf{v}|\mathbf{u}_0) F(\mathbf{v}')}{\sum_{V'_\mathbf{v}} F(\mathbf{v}')} \end{aligned} \quad (129)$$

where (a) follows from Lemma 10, and (b) follows because $F(\mathbf{v}')$ defined in (117) depends only on the composition of \mathbf{v}' and is same for all sequences in $V'_\mathbf{v}$.

We will first bound the denominator in (129). Let $f(y'|x)$ denote a conditional probability distribution with $f(y') = \sum_{\mathcal{X}} P(x) f(y'|x)$. This defines $F(\mathbf{v}'|\mathbf{u})$ obtained as the product of the values of $f(y'|x)$ for each corresponding pair of events of the sequences \mathbf{v}' and \mathbf{u} . If \mathbf{u} is any sequence consisting of letters $x \in \mathcal{X}$, then we define $P(\mathbf{u}) = \prod_{\mathcal{X}} P(x)^{n(x)}$ where $n(x)$ is the number of occurrences of letter x in \mathbf{u} . If U is the space of all possible length n sequences

consisting of letters $x \in \mathcal{X}$, and U_0 denotes the subset of sequences having the given subblock-composition, then

$$F(\mathbf{v}') = \sum_U P(\mathbf{u}) F(\mathbf{v}'|\mathbf{u}) \geq \sum_{U_0} P(\mathbf{u}) F(\mathbf{v}'|\mathbf{u}) \quad (130)$$

It follows that

$$\sum_{V'_{\mathbf{v}}} F(\mathbf{v}') \geq \sum_{U_0} P(\mathbf{u}) \sum_{V'_{\mathbf{v}}} F(\mathbf{v}'|\mathbf{u}) \quad (131)$$

$$\stackrel{(c)}{=} \left(\sum_{U_0} P(\mathbf{u}) \right) \left(\sum_{V'_{\mathbf{v}}} F(\mathbf{v}'|\mathbf{u}) \right) \quad (132)$$

where (c) follows because $\sum_{V'_{\mathbf{v}}} F(\mathbf{v}'|\mathbf{u})$ is same for all $\mathbf{u} \in U_0$. Further,

$$\begin{aligned} \sum_{U_0} P(\mathbf{u}) &= \left(\frac{L!}{\prod_{\mathcal{X}} (L P(x))!} \right)^m \left(\prod_{\mathcal{X}} P(x)^{L P(x)} \right)^m \\ &= \exp(-n r(L, P)) \end{aligned} \quad (133)$$

with $r(L, P)$ given by (30). Thus, we have

$$\sum_{V'_{\mathbf{v}}} F(\mathbf{v}') \geq \exp(-n r(L, P)) \sum_{V'_{\mathbf{v}}} F(\mathbf{v}'|\mathbf{u}). \quad (134)$$

We will now bound the numerator in (129). For each $x \in \mathcal{X}$, define the logarithm of the moment-generating function

$$\begin{aligned} \gamma_x(w_1, w_2) &:= \log \sum_{y \in \mathcal{Y}} \sum_{y' \in \mathcal{Y}} \exp(w_1 \tilde{D}(x, y) + \\ &\quad w_2 [\tilde{D}(x, y') - \tilde{D}(x, y)]) f(y') W(y|x) \end{aligned} \quad (135)$$

where w_1 and w_2 are parameters associated with the random variables $\tilde{D}(x, y)$ and $\tilde{D}(x, y') - \tilde{D}(x, y)$, respectively. Define the tilted probability distributions

$$\begin{aligned} Q_0(y, y'|x) &:= \exp(w_1 \tilde{D}(x, y) + w_2 [\tilde{D}(x, y') - \tilde{D}(x, y)]) \\ &\quad - \gamma_x(w_1, w_2) f(y') W(y|x) \end{aligned} \quad (136)$$

$$\begin{aligned} Q_0(\mathbf{v}, \mathbf{v}'|\mathbf{u}_0) &:= \exp(w_1 \tilde{D}(\mathbf{u}_0, \mathbf{v}) + w_2 [\tilde{D}(\mathbf{u}_0, \mathbf{v}') - \tilde{D}(\mathbf{u}_0, \mathbf{v})]) \\ &\quad - n \gamma_0(w_1, w_2) F(\mathbf{v}') W^n(\mathbf{v}|\mathbf{u}_0) \end{aligned} \quad (137)$$

where $\gamma_0(w_1, w_2) = \sum_{\mathcal{X}} P(x) \gamma_x(w_1, w_2)$. From (137), it follows that

$$\begin{aligned} \sum_{V'_{0\mathbf{v}}} W^n(\mathbf{v}|\mathbf{u}_0) F(\mathbf{v}') &\leq \exp(n\gamma_0(w_1, w_2) - w_1 D(\mathbf{u}_0, \mathbf{v})) \\ &\times \sum_{V'_{0\mathbf{v}}} Q_0(\mathbf{v}, \mathbf{v}'|\mathbf{u}_0); \quad w_2 \leq 0 \end{aligned} \quad (138)$$

The RHS of (138) is minimized by setting $w_1 = 2w_2 + 1$, in which case we have

$$Q_0(y, y'|x) = Q_0(y|x) Q_0(y'|x) \quad (139)$$

$$Q_0(y|x) = \frac{W(y|x)^{(1-w_1)/2} f(y)^{(1+w_1)/2}}{\sum_y W(y|x)^{(1-w_1)/2} f(y)^{(1+w_1)/2}} \quad (140)$$

$$Q_0(\mathbf{v}, \mathbf{v}'|\mathbf{u}_0) = Q_0(\mathbf{v}|\mathbf{u}_0) Q_0(\mathbf{v}'|\mathbf{u}_0) \quad (141)$$

Using (138), (141), and the fact that $V'_{0\mathbf{v}} \subset V'_{\mathbf{v}}$, we have

$$\begin{aligned} \sum_{V'_{0\mathbf{v}}} W^n(\mathbf{v}|\mathbf{u}_0) F(\mathbf{v}') &\leq \exp(n\gamma_0(w_1, w_2) - w_1 \tilde{D}(\mathbf{u}_0, \mathbf{v})) \\ &\times Q_0(\mathbf{v}|\mathbf{u}_0) \sum_{V'_{\mathbf{v}}} Q_0(\mathbf{v}'|\mathbf{u}_0); \quad w_1 \leq 1 \end{aligned} \quad (142)$$

If we let $f(y'|x) = Q_0(y'|x)$ then $F(\mathbf{v}') = Q_0(\mathbf{v}'|\mathbf{u}_0)$ and P_1 can be bounded using (129), (134), and (142) as

$$\begin{aligned} P_1 &\leq \exp(nr(L, P) + n\gamma_0(w_1, w_2)) \times \\ &\sum_{V_0} \exp(-w_1 \tilde{D}(\mathbf{u}_0, \mathbf{v})) Q_0(\mathbf{v}|\mathbf{u}_0); \quad w_1 \leq 1 \end{aligned}$$

Since all the sequence \mathbf{v} belonging to V_0 have a distance from \mathbf{u}_0 which does not exceed D_0 , we have

$$P_1 \leq \exp(nr(L, P) + n\gamma_0(w_1, w_2) - w_1 D_0); \quad w_1 \leq 0 \quad (143)$$

From (128), we know that $D_0 = n\gamma'(s)$, and RHS of (143) is minimized when $w_1 = 2s - 1$, $w_2 = s - 1$, and s satisfies $0 \leq s \leq 1/2$. In this case, we have

$$Q_0(y|x) = V(y|x), \quad (144)$$

$$\gamma_0(w_1, w_2) = 2\gamma(s), \quad (145)$$

where $V(y|x)$ is given by (127). Now (143) takes the form

$$P_1 \leq \exp(-n[(2s - 1)\gamma'(s) - 2\gamma(s) - r(L, P)]); \quad 0 \leq s \leq 1/2 \quad (146)$$

Note that since $f(y|x) = Q_0(y|x)$, it follows from (144) that

$$f(y) = PV(y) . \quad (147)$$

Since $R = \frac{\log M}{n}$, and $P_e < MP_1 + P_2$, we have

$$\begin{aligned} P_e &< \exp(-n[(2s-1)\gamma'(s) - 2\gamma(s) - (R + r(L, P))]) \\ &\quad + \exp(-n[s\gamma'(s) - \gamma(s)]) ; \quad 0 \leq s \leq 1/2. \end{aligned} \quad (148)$$

Now, if we choose s such that

$$(s-1)\gamma'(s) - \gamma(s) = R + r(L, P) ; \quad 0 \leq s \leq 1/2 , \quad (149)$$

then it follows from (148), (149), and (125) that

$$P_e < 2 \exp(-nD(V||W|P)) . \quad (150)$$

From (127) and (147) we have

$$V(y|x) = \frac{W(y|x)^{1-s}PV(y)^s}{\sum_y W(y|x)^{1-s}PV(y)^s} . \quad (151)$$

Now $(s-1)\gamma'(s) - \gamma(s)$ is a decreasing function of s for $0 \leq s \leq 1/2$ (its derivative is $(s-1)\gamma''(s)$). Let \hat{R} denote its value at $s = 1/2$ (that is $\hat{R} = -0.5\gamma'(0.5) - \gamma(0.5)$). From (125), (128), and (147), it follows that condition (149) can be equivalently be expressed as

$$I(P, V) = R + r(L, P) , \quad \text{if } R + r(L, P) \geq \hat{R}. \quad (152)$$

If we let

$$R' = R + r(L, P) , \quad (153)$$

then using Lemma 2, (151), and (152), we observe that $D(V||W|P) = E_{sp}(R', P, W)$, and (150) is equivalent to

$$P_e < 2 \exp(-nE_{sp}(R', P, W)) ; \quad \text{if } R' \geq \hat{R}. \quad (154)$$

Note that for $0 \leq s \leq 1/2$, we have $E_{sp}(R', P, W) = s\gamma'(s) - \gamma(s)$ and $R' = (s-1)\gamma'(s) - \gamma(s)$.

Thus

$$\frac{d}{dR'} E_{sp}(R', P, W) = -\frac{s}{1-s} ; \quad 0 \leq s \leq 1/2 \quad (155)$$

and the slope of $E_{sp}(R', P, W)$ with respect to R' at $s = 1/2$ (corresponding to $R' = \hat{R}$) is -1 .

For obtaining a bound on P_e when $R' < \hat{R}$, we let radius $D_0 = \infty$. In this case,

$$P_1 = \sum_{\tilde{V}} P(\mathbf{v}|\mathbf{u}_0) Pr[\tilde{D}(\mathbf{u}', \mathbf{v}) \leq \tilde{D}(\mathbf{u}_0, \mathbf{v})] \quad (156)$$

where \tilde{V} is the space of all possible output sequences. Note that $P_2 = 0$ when $D_0 = \infty$. The upper bound for P_1 in (156) can be calculated the same way as before. However, the condition $D(\mathbf{u}_0, \mathbf{v}) \leq D_0$ is eliminated by setting $w_1 = 0$ in (135). Here, the upper bound for P_1 is obtained by setting $w_2 = s - 1$ and $s = 1/2$. The probability of error in this case is bounded as

$$\begin{aligned} P_e &< \exp(-n(-2\gamma(0.5) - R')) ; \text{ if } R' < \hat{R} \\ &= \exp\left(-n\left(E_{sp}(\hat{R}, P, W) + \hat{R} - R'\right)\right) ; \text{ if } R' < \hat{R} \end{aligned} \quad (157)$$

and the proof is complete by combining (153), (154) and (157).

APPENDIX H

PROOF OF THEOREM 8

For the proof, we will construct a simple example of a symmetric DMC for which the uniform distribution over \mathcal{A} does not achieve SECC capacity.

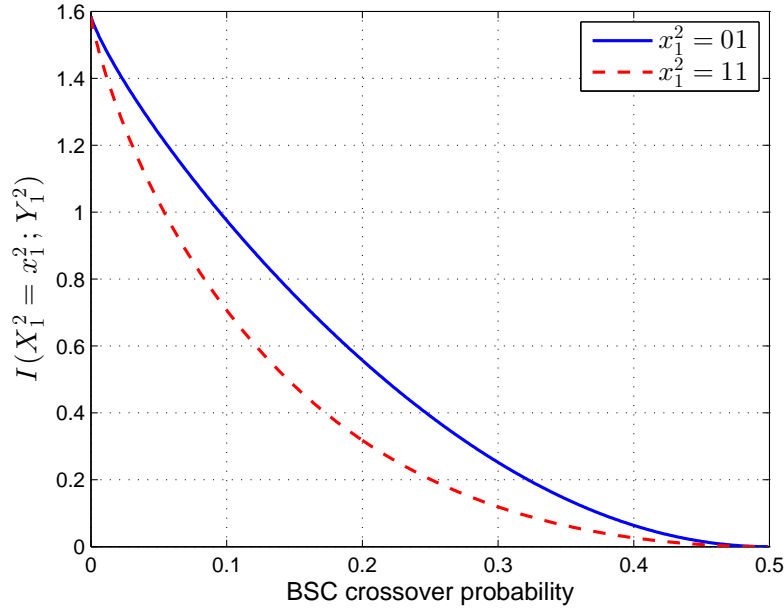


Fig. 9. $I(X_1^L = x_1^L; Y_1^L)$ versus BSC crossover probability p_0 for $L = 2$.

Consider the following parameters for a BSC with crossover probability p_0 :

$$b(0) = 0, \quad b(1) = 1, \quad B = 0.5, \quad L = 2, \quad 0 < p_0 < 0.5. \quad (158)$$

With the above parameters, the input alphabet for the induced vector channel is given by $\mathcal{A} = \{01, 10, 11\}$. A uniform distribution over \mathcal{A} will achieve SECC capacity if and only if $I(X_1^L = x_1^L; Y_1^L)$ is same for all $x_1^L \in \mathcal{A}$ [60, Thm. 4.5.1], where

$$I(X_1^L = x_1^L; Y_1^L) = \sum_{y_1^L \in \mathcal{Y}^L} W^L(y_1^L | x_1^L) \log \frac{|\mathcal{A}| W^L(y_1^L | x_1^L)}{\sum_{\tilde{x}_1^L \in \mathcal{A}} W^L(y_1^L | \tilde{x}_1^L)}. \quad (159)$$

The proof is completed by numerically verifying that for BSC having parameters given by (158), $I(X_1^L = 01; Y_1^L) \neq I(X_1^L = 11; Y_1^L)$. Fig. 9 shows that $I(X_1^L = 01; Y_1^L)$ and $I(X_1^L = 11; Y_1^L)$ are different when $0 < p_0 < 0.5$. ■

REFERENCES

- [1] A. Tandon, M. Motani, and L. R. Varshney, “Constant subblock composition codes for simultaneous energy and information transfer,” in *Proc. IEEE SECON 2014 Workshop on Energy Harvesting Communications*, Jun. 2014, pp. 45–50.
- [2] —, “Real-time simultaneous energy and information transfer,” to appear in *Proc. 2015 IEEE Int. Symp. Inf. Theory*.
- [3] S. Hui, “Planar wireless charging technology for portable electronic products and Qi,” *Proceedings of the IEEE*, vol. 101, no. 6, pp. 1290–1301, Jun. 2013.
- [4] A. Yakovlev, S. Kim, and A. Poon, “Implantable biomedical devices: Wireless powering and communication,” *IEEE Commun. Mag.*, vol. 50, no. 4, pp. 152–159, Apr. 2012.
- [5] W. C. Brown, “The history of power transmission by radio waves,” *IEEE Trans. Microw. Theory Tech.*, vol. 32, no. 9, pp. 1230–1242, Sept. 1984.
- [6] M. Radovsky, *Alexander Popov: Inventor of the Radio*. Moscow: Foreign Language Publishing House, 1957.
- [7] D. T. Emerson, “The work of Jagadis Chandra Bose: 100 years of millimeter-wave research,” *IEEE Trans. Microw. Theory Tech.*, vol. 45, no. 12, pp. 2267–2273, Dec 1997.
- [8] G. Marconi, “Wireless telegraphic communication,” *Nobel Lectures*, Dec. 1909.
- [9] D. D. Hoolihan, “The Titanic and radio frequency interference,” *IEEE Electromagn. Compat. Mag.*, vol. 1, no. 2, July 2012.
- [10] L. R. Varshney, “Transporting information and energy simultaneously,” in *Proc. 2008 IEEE Int. Symp. Inf. Theory*, Jul. 2008, pp. 1612–1616.
- [11] P. Grover and A. Sahai, “Shannon meets Tesla: Wireless information and power transfer,” in *Proc. 2010 IEEE Int. Symp. Inf. Theory*, Jun. 2010, pp. 2363–2367.
- [12] P. Popovski, A. M. Fouladgar, and O. Simeone, “Interactive joint transfer of energy and information,” *IEEE Trans. Commun.*, vol. 61, no. 5, pp. 2086–2097, May 2013.
- [13] Z. Ding, C. Zhong, D. Wing Kwan Ng, M. Peng, H. A. Suraweera, R. Schober, and H. V. Poor, “Application of smart antenna technologies in simultaneous wireless information and power transfer,” *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 86–93, Apr. 2015.
- [14] S. Timotheou, I. Krikidis, S. Karachontzitis, and K. Berberidis, “Spatial domain simultaneous information and power transfer for MIMO channels,” accepted for publication in *IEEE Trans. Wireless Commun.*, 2015.
- [15] R. Zhang, L.-L. Yang, and L. Hanzo, “Energy pattern aided simultaneous wireless information and power transfer,” accepted for publication in *IEEE J. Sel. Areas Commun.*, 2015.

- [16] R. R. Harrison, P. T. Watkins, R. J. Kier, R. O. Lovejoy, D. J. Black, B. Greger, and F. Solzbacher, "A low-power integrated circuit for a wireless 100-electrode neural recording system," *IEEE J. Solid-State Circuits*, vol. 42, no. 1, pp. 123–133, 2007.
- [17] G. Yilmaz, O. Atasoy, and C. Dehollain, "Wireless energy and data transfer for in-vivo epileptic focus localization," *IEEE Sensors J.*, vol. 13, no. 11, pp. 4172–4179, Nov. 2013.
- [18] Á. I. Barbero, E. Rosnes, G. Yang, and Ø. Ytrehus, "Constrained codes for passive RFID communication," in *Proc. 2011 Inf. Theory Appl. Workshop*, Feb. 2011.
- [19] A. M. Fouladgar, O. Simeone, and E. Erkip, "Constrained codes for joint energy and information transfer," *IEEE Trans. Commun.*, vol. 62, no. 6, pp. 2121–2131, Jun. 2014.
- [20] A. Tandon, M. Motani, and L. R. Varshney, "On code design for simultaneous energy and information transfer," in *Proc. 2014 Inf. Theory Appl. Workshop*, Feb. 2014.
- [21] K. A. S. Immink, P. H. Siegel, and J. K. Wolf, "Codes for digital recorders," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2260–2299, Oct. 1998.
- [22] K. A. S. Immink, "Runlength-limited sequences," *Proc. IEEE*, vol. 78, no. 11, pp. 1745–1759, Nov. 1990.
- [23] R. D. Cideciyan, F. Dolivo, R. Hermann, W. Hirt, and W. Schott, "A PRML system for digital magnetic recording," *IEEE J. Sel. Areas Commun.*, vol. 10, no. 1, pp. 38–56, Jan. 1992.
- [24] E. Zehavi and J. K. Wolf, "On runlength codes," *IEEE Trans. Inf. Theory*, vol. 34, no. 1, pp. 45–54, 1988.
- [25] S. Shamai and Y. Kofman, "On the capacity of binary and Gaussian channels with run-length-limited inputs," *IEEE Trans. Commun.*, vol. 38, no. 5, pp. 584–594, 1990.
- [26] P. Jacquet and W. Szpankowski, "Noisy constrained capacity for BSC channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5412–5423, Nov. 2010.
- [27] Y. M. Chee, Z. Cherif, J.-L. Danger, S. Guilley, H. M. Kiah, J.-L. Kim, P. Sole, and X. Zhang, "Multiply constant-weight codes and the reliability of loop physically unclonable functions," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7026–7034, Nov. 2014.
- [28] Z. Cherif, J.-L. Danger, S. Guilley, J.-L. Kim, and P. Sole, "Multiply constant weight codes," in *Proc. 2013 IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 306–310.
- [29] S. Johnson, "Upper bounds for constant weight error correcting codes," *Discrete Math.*, vol. 3, no. 1-3, pp. 109–124, 1972.
- [30] E. Ordentlich and R. M. Roth, "Two-dimensional weight-constrained codes through enumeration bounds," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1292–1301, Jul. 2000.
- [31] Y. M. Chee, H. M. Kiah, and P. Purkayastha, "Matrix codes and multitone frequency shift keying for power line communications," in *Proc. 2013 IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 2870–2874.
- [32] N. Pavlidou, A. J. H. Vinck, J. Yazdani, and B. Honary, "Power line communications: state of the art and future trends," *IEEE Commun. Mag.*, vol. 41, no. 4, pp. 34–40, Apr. 2003.
- [33] W. Chu, C. J. Colbourn, and P. Dukes, "Constructions for permutation codes in powerline communications," *Des. Codes Cryptog.*, vol. 32, pp. 51–64, 2004.
- [34] —, "On constant composition codes," *Discrete Appl. Math.*, vol. 154, no. 6, pp. 912–929, Apr. 2006.
- [35] Y. M. Chee, H. M. Kiah, P. Purkayastha, and C. Wang, "Importance of symbol equity in coded modulation for power line communications," in *Proc. 2012 IEEE Int. Symp. Inf. Theory*, Jul. 2012, pp. 661–665.
- [36] O. Ozel and S. Ulukus, "AWGN channel under time-varying amplitude constraints with causal information at the transmitter," in *Conf. Rec. 45th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2011, pp. 373–377.
- [37] —, "Achieving AWGN capacity under stochastic energy harvesting," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6471–6483, Oct. 2012.

- [38] Y. Dong and A. Ozgur, "Approximate capacity of energy harvesting communication with finite battery," in *Proc. 2014 IEEE Int. Symp. Inf. Theory*, June 2014, pp. 801–805.
- [39] V. Jog and V. Anantharam, "An energy harvesting AWGN channel with a finite battery," in *Proc. 2014 IEEE Int. Symp. Inf. Theory*, Jun. 2014, pp. 806–810.
- [40] R. Rajesh, V. Sharma, and P. Viswanath, "Capacity of Gaussian channels with energy harvesting and processing cost," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2563–2575, May 2014.
- [41] W. Mao and B. Hassibi, "On the capacity of a communication system with energy harvesting and a limited battery," in *Proc. 2013 IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 1789–1793.
- [42] S. Ulukus, A. Yener, E. Erkip, O. Simeone, M. Zorzi, P. Grover, and K. Huang, "Energy harvesting wireless communications: A review of recent advances," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 3, pp. 360–381, Mar. 2015.
- [43] L. R. Varshney, "Unreliable and resource-constrained decoding," Ph.D. thesis, Massachusetts Institute of Technology, Jun. 2010.
- [44] X. Zhou, R. Zhang, and C. K. Ho, "Wireless information and power transfer: Architecture design and rate-energy tradeoff," *IEEE Trans. Commun.*, vol. 61, no. 11, pp. 4754–4767, Nov. 2013.
- [45] R. M. Fano, *Transmission of Information: A Statistical Theory of Communications*. Cambridge, MA: MIT Press, 1961.
- [46] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems (2nd ed)*. Cambridge Univ. Press, 2011.
- [47] A. D. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications–I," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 6, pp. 769–772, Nov. 1973.
- [48] N. Chayat and S. Shamai, "Extension of an entropy property for binary input memoryless symmetric channels," *IEEE Trans. Inf. Theory*, vol. 35, no. 5, pp. 1077–1079, Sept. 1989.
- [49] R. Ahlswede and J. Körner, "On the connection between the entropies of input and output distributions of discrete memoryless channels," in *Proc. 5th Conf. Probability Theory, Brasov 1974*. Ed. Acad. Republicii Socialiste România, 1977, pp. 13–22.
- [50] S. Yekhanin, "Locally decodable codes," *Found. Trends Theor. Comput. Sci.*, vol. 7, no. 1, pp. 1–117, 2011.
- [51] Y. Polyanskiy, "Channel coding: Non-asymptotic fundamental limits," Ph.D. thesis, Princeton University, Nov. 2010.
- [52] P. Moulin, "The log-volume of optimal constant-composition codes for memoryless channels, within $O(1)$ bits," in *Proc. 2012 IEEE Int. Symp. Inf. Theory*, Jul. 2012, pp. 826–830.
- [53] M. Tomamichel and V. Tan, "A tight upper bound for the third-order asymptotics for most discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7041–7051, Nov. 2013.
- [54] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
- [55] L. Weiss, "On the strong converse of the coding theorem for symmetric channels without memory," *Q. Appl. Math.*, vol. 18, no. 3, pp. 209–214, Oct. 1960.
- [56] V. Strassen, "Asymptotische abschätzungen in Shannon's informationstheorie," in *Trans. 3rd Prague Conf. Inf. Theory*, Prague, 1962, pp. 689–723.
- [57] L. R. Varshney, "On energy/information cross-layer architectures," in *Proc. 2012 IEEE Int. Symp. Inf. Theory*, Jul. 2012, pp. 1361–1365.
- [58] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 18, no. 1, pp. 14–20, Jan. 1972.
- [59] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 460–473, Jul. 1972.

- [60] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley and Sons, Inc., 1968.
- [61] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY: Cambridge University Press, 2004.